(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 December 2004 (16.12.2004)

PCT

(10) International Publication Number
WO 2004/108897 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US2004/017552

(22) International Filing Date: 2 June 2004 (02.06.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/475,576 2 June 2003 (02.06.2003) US
60/520,044 13 November 2003 (13.11.2003) US

(71) Applicant *(for all designated States except US)*: **CYTOKINETICS, INC.** [US/US]; 280 East Grand Avenue, South San Francisco, CA 94080 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BERAUD, Christophe** [FR/US]; 649 North Farrell Drive, Palm Springs, CA 92262 (US). **MIRZA, Amer, Mushtaq** [CA/US]; 271 6th Avenue Apartment #13, San Francisco, CA 94118 (US).

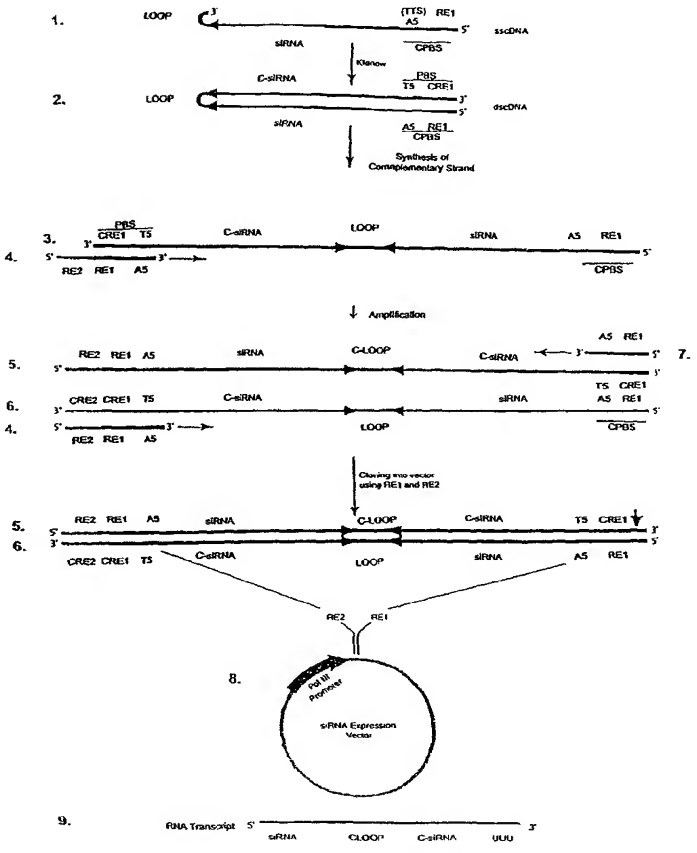
(74) Agents: **LIEBESCHUETZ, Joe** et al.; Two Embarcadero Center 8th Floor, San Francisco, CA 94111-3834 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH,

[Continued on next page]

(54) Title: SIRNA LIBRARIES



(57) Abstract: The invention provides methods of expressing an siRNA starting from a single template encoding one strand of the siRNA. The methods are particularly suitable for generating large libraries of random siRNAs in which one strand of each siRNA can be synthesized simultaneously in a solid-phase synthesis procedure. Methods for normalizing siRNA libraries to enrich for siRNAs of interest and methods of expressing siRNAs are also provided.

WO 2004/108897 A2

GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

siRNA Libraries

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] The present application is a non-provisional of and claims the benefit of USSN 60/475,576, filed on June 2, 2003, and is a non-provisional of and claims the benefit of USSN 60/520,044, filed on November 13, 2003, both of which are hereby incorporated by reference in their entirety for all purposes.

FIELD OF THE INVENTION

[0002] The invention resides in the fields of molecular genetics and functional genomics.

BACKGROUND OF THE INVENTION

[0003] A number of technologies have been developed for the regulation of gene function. These technologies include use of anti-sense RNAs, Milner et al., 1997, Nature Biotechnology 15:537), zinc finger proteins (WO 00/00409), ribozymes (Haseloff & Gerlach, (1988) Nature 334:585-591 and Uhlenbeck, (1987) Nature 328:596-603 and US 5,496,698) and more recently small interfering RNA (siRNA). siRNAs are relatively short, at least partly double stranded, RNA molecules that serve to inhibit expression of a complementary mRNA transcript. It has been proposed that siRNAs act by inducing endonuclease-mediated degradation of a complementary mRNA transcript. Principles for design and use of siRNAs generally are described by WO 99/32619, Elbashir, EMBO J. 20, 6877-6888 (2001) and Nykanen et al., Cell 107, 309-321 (2001); WO 01/29058.

[0004] Although siRNAs have shown much promise as agents for suppressing regulation of genes the delivery of an siRNA to a cell presents certain difficulties in view of the double stranded nature of siRNAs and the fact that RNAs are not naturally transcribed in a double stranded format. One means of delivery is to prepare siRNAs in vitro and deliver siRNAs directly rather than in the form of DNA encoding an siRNA. However, such an approach is limited in the intracellular level and persistence of siRNA that can be achieved.

[0005] siRNAs can also be synthesized within a cell by preparing a vector in which a segment of DNA encoding the siRNA is inserted flanking a pair of promoters that are oriented to drive transcription of the inserted segments in opposite orientations. Transcription

from such promoters produces two complementary RNA strands that can subsequently anneal to form the desired dsRNA. Alternatively, DNA segments encoding the strands of the siRNA are separately inserted downstream of a single promoter. In this system, the sense and antisense strands of the siRNA are cotranscribed to generate a single RNA strand that is self-complementary and thus can form dsRNA.

[0006] Production of random libraries of siRNAs presents special challenges in obtaining large numbers of double stranded RNA molecules without individually synthesizing each one. For example, it is relatively simple to generate a random library of single-stranded RNA by using randomized mixes of ribonucleotides in a solid phase synthesis procedure. However, the difficulty there lies in how each single stranded random RNA molecule is to be associated with its complementary strand to form an siRNA.

SUMMARY OF THE CLAIMED INVENTION

[0007] Provided herein are methods of preparing a vector for expression of an siRNA, comprising providing a first DNA molecule comprising a segment encoding a strand of an siRNA and a loop segment at the 3' end, extending the 3' end of the loop with a polymerase to form a second DNA molecule duplexed with the first DNA molecule, the second DNA molecule encoding the complementary strand of the siRNA, denaturing the duplexed first and second DNA molecules to form a single-stranded DNA molecule comprising the second DNA molecule fused to the 3' end of the first DNA molecule, synthesizing a second strand complementary to the single-stranded DNA molecule forming a double stranded DNA molecule, and cloning at least a fragment of the double stranded molecule into a vector in operable linkage with a promoter whereby transcription of the vector forms a transcript comprising a segment having the siRNA strand sequence, the loop segment or its complement, and the complement of the siRNA strand, and the transcript self-anneals to form an siRNA. In some methods the providing step is performed by synthesizing the first DNA molecule using solid-phase synthesis. In some methods the first DNA molecule of the providing step is formed by ligation of the loop segment to the 3' end of the segment encoding a strand of an siRNA.

[0008] Some methods further comprise amplifying the double stranded DNA molecule of the synthesizing step. In an embodiment of this synthesizing step, the primer binding to the primer binding site on the second DNA molecule can further comprise a 5' segment

comprising a strand of a type IIs restriction endonuclease recognition site whereby digestion of the double stranded DNA molecule with the type IIs restriction endonuclease removes the complement of the transcription termination segment from the second DNA molecule.

[0009] In some methods the first DNA molecule further comprises the complement of a primer binding site at the 5' end, the second DNA molecule comprises the primer binding site at its 3' end, and the synthesizing step is performed using a primer binding to the primer binding site. The first DNA molecule can further comprise a transcription termination segment optionally within the complement of the primer binding site, and the second DNA molecule can further comprise the complement of the transcription termination segment. The first DNA molecule can also further comprise a strand of a restriction enzyme recognition site optionally within the complement of the primer binding site.

[0010] In some methods the siRNA strand has a random sequence. In some methods a plurality of first DNA molecules encoding different siRNA strand sequences are synthesized, each of the different sequences being a random sequence, and the method forms a plurality of vectors encoding different siRNAs. Some methods further comprise transforming the plurality of vectors into a population of cells and identifying which cells develop a change in a property responsive to suppression of expression of a gene by an siRNA.

[0011] In some methods the siRNA strands are a normalized population selected from a random population by hybridization to a nucleic acid library. In some methods the nucleic acid library is an mRNA library or a cDNA library. Some methods further comprise contacting the DNA segments encoding the siRNA strands with a nucleic acid library under hybridization conditions, and selecting the DNA segments that hybridize to the nucleic acid library. Some methods further comprise linking the DNA segments that hybridize to the nucleic acid library to the loop segment to provide the plurality of first DNA molecules. In one embodiment of these methods, members of the library are conjugated to an immobilizable tag and the selecting comprises immobilizing the nucleic acid library to a solid support; contacting the DNA segments with the immobilized nucleic acid library under hybridization conditions; washing the immobilized nucleic acid library to remove unhybridized or weakly hybridized DNA segments; and eluting the hybridized DNA segments. In some methods the immobilizable tag is biotin. In some methods the washing comprises adding progressively lower salt concentrations of wash buffer to the immobilized nucleic acid library. In some methods the salt is NaCl.

[0012] In some methods the providing step comprises providing an RNA molecule comprising a segment encoding the complementary strand of the siRNA strand sequence flanked at its 5' end by an RNA segment having the complement of the loop sequence, and at its 3' end by an RNA segment having the complement of the primer binding site sequence, extending a primer bound to the RNA segment having the complement of the primer binding site sequence with reverse polymerase to generate the first DNA molecule complexed with the RNA molecule, and digesting the RNA molecule to leave the first DNA molecule free of the RNA molecule.

[0013] Also disclosed are methods of enriching a random nucleic acid library, comprising hybridizing the random nucleic acid library to an mRNA or cDNA population; and isolating members of the random nucleic acid library that bind to the population. In some methods each member of the mRNA or cDNA population is linked to an immobilizable tag. In some methods the immobilizable tag is biotin. In some methods the random nucleic acid library is a library of DNA segments encoding random RNA sequences. In some methods the library of DNA segments is linked to an immobilizable tag. In some methods the immobilizable tag is biotin. In some methods the DNA segments are linked to the immobilizable tag via a nuclease-resistant DNA segment, a disulphide bond, and a nuclease resistant linker. Some methods further comprise contacting hybrids of the DNA segments and mRNA or cDNA population with a single-stranded nuclease to digest unhybridized DNA segments and unhybridized members of the mRNA or cDNA population. Additional methods further comprise subjecting the hybrids to denaturing conditions to separate the DNA segments from the members of the mRNA or cDNA population to which the DNA segments hybridized. Other methods further comprise immobilizing the DNA segments via the immobilizable tag. Some methods further comprise cleaving the disulfide bond to separate the DNA segments from the immobilizable tag.

[0014] Provided herein are DNA constructs comprising a Pol III promoter; a segment encoding an RNA molecule; a transcriptional termination segment; and at least one regulatory element, wherein the promoter, the segment encoding an RNA molecule, the termination segment and the at least one regulatory element are in operable linkage such that a stimulus can act via the at least one regulatory element to induce the promoter to initiate transcription of the segment encoding an RNA molecule, transcription being terminated by the transcriptional termination segment. In some constructs the RNA molecule is an siRNA. In some constructs the promoter contains a tet operator segment. In some constructs the promoter is a type III Pol III promoter. In some constructs a tet operator segment is located

in the promoter. In some constructs the tet operator segment is tet O1. In some constructs the tet operator segment is tet O2. In some constructs the transcriptional termination segment is a U6 Pol III terminator.

[0015] Provided herein are RNA expression vectors, wherein each vector contains: a Pol III promoter; a segment encoding an RNA molecule, the RNA molecule sequence differing between vectors; a transcriptional termination segment; and at least one regulatory element; wherein the promoter, the segment encoding an RNA molecule, the termination segment and the at least one regulatory element are in operable linkage such that a stimulus can act via the at least one regulatory element to induce the promoter to initiate transcription of the segment encoding an RNA molecule, transcription being terminated by the transcriptional termination segment. In some libraries each promoter contains a tet operator segment. In some libraries each promoter is a U6 promoter.

[0016] Provided herein are methods of expressing an RNA molecule comprising providing a cell containing a DNA construct comprising a Pol III promoter, a segment encoding an RNA molecule, a transcriptional termination segment, and at least one regulatory element, wherein the promoter, the segment encoding an RNA molecule, the termination segment and the at least one regulatory element are in operable linkage such that a stimulus can act via the at least one regulatory element to induce the promoter to initiate transcription of the segment encoding an RNA molecule, transcription being terminated by the transcriptional termination segment; and providing the stimulus, which acts via the regulatory element to induce the Pol III promoter. In some methods the promoter contains a tet operator segment, the cell expresses the tet repressor, and the stimulus is provided by removing tetracycline from the culture media.

[0017] Provided herein are methods of identifying a gene comprising synthesizing a library of random RNA segments; cloning the segments into a population of DNA constructs containing a Pol III promoter, a segment encoding an RNA molecule, a transcriptional termination segment, and at least one regulatory element, wherein the promoter, the segment encoding an RNA molecule, the termination segment and the at least one regulatory element are in operable linkage such that a stimulus can act via the at least one regulatory element to induce the promoter to initiate transcription of the segment encoding an RNA molecule, transcription being terminated by the transcriptional termination segment; transforming a population of cells with the population of DNA constructs containing the random RNA segments; providing a stimulus that acts via the at least one regulatory element to induce the promoter; and screening or selecting for a desired phenotype. In some methods the promoter

contains a tet operator segment. In some methods the promoter is the U6 promoter. In some methods the promoter is the U6 promoter and a tet operator segment is located between the proximal sequencing element and the TATA box. Some methods further comprise the steps of recovering a DNA construct from a cell which exhibits the desired phenotype; and sequencing the RNA segment contained in the construct. In some methods the RNA segments are siRNA.

BRIEF DESCRIPTION OF THE FIGURES

[0018] Figure 1 shows steps of an exemplary a method of synthesizing a library of siRNA molecules using a cDNA as a starting molecule.

[0019] Figure 2 shows steps of an exemplary a method of synthesizing a library of siRNA molecules using a cDNA as a starting molecule.

[0020] Figures 3A-3D show a method of normalizing a siRNA library.

[0021] Figure 4 shows a construct for expressing a green fluorescent protein gene using an inducible Tet O2 sequence in the promoter.

[0022] Figures 5A and 5B show cells transfected with an inducible GFP expression vector. Panel A shows fluorescence and panel B shows all cells in the fields of the A panels.

[0023] Figure 6 shows relative GFP expression in transfected cells using an inducible promoter.

[0024] Figure 7 shows a U6 promoter cassette.

[0025] Figure 8 shows the sequence of an inducible Pol III promoter construct.

[0026] Figure 9 shows an inducible U6-siRNA cassette construct.

[0027] Figure 10 shows relative expression of a nucleic acid segment in cells transfected with the construct.

[0028] Figure 11 shows steps 1-3 of an exemplary method of normalizing an siRNA library.

[0029] Figure 12 shows steps 4-7 of an exemplary method of normalizing an siRNA library.

[0030] Figure 13 shows elution of specific oligonucleotides from an immobilized library of nucleic acid described in Example 3.

DEFINITIONS

[0031] siRNAs have a double stranded region and optionally single stranded overhangs. For such siRNAs, sequence identity between the siRNA and a target sequence in a transcript to be regulated is determined by comparing the double stranded region of the siRNA with the target sequence.

[0032] The terms "modulating expression" "inhibiting expression" " and "silencing" refer to the ability of an siRNA to inhibit translation of a gene.

[0033] An siRNA inhibits expression of a transcript or protein if the level of transcript of protein is reduced by at least 10% and preferably to an undetectable extent compared with a control cell or organism in which the siRNA inhibitor is absent.

[0034] A strand of an siRNA having the same polarity as a transcript is referred to as the sense strand, and the other strand of an siRNA as the antisense strand.

[0035] A target site is a segment of nucleotides in a polynucleotide to be bound by the siRNA. Not every nucleotide in the target site necessarily matches the siRNA sequence.

[0036] Unless otherwise apparent from the context, reference to an RNA inhibitor, includes the RNA itself and DNA encoding the RNA.

[0037] The term "patient" includes human and other mammalian subjects that receive either prophylactic or therapeutic treatment.

[0038] The term "naked polynucleotide" refers to a polynucleotide not complexed with colloidal materials. Naked polynucleotides are sometimes cloned in a plasmid vector.

[0039] Unless otherwise apparent from the context, the term "random" includes fully random and partially random segments.

[0040] Compositions or methods "comprising" one or more recited elements may include other elements not specifically recited. For example, a composition that comprises a siRNA expression vector both the siRNA expression vector and transformation reagents or a pharmaceutical carrier.

[0041] In some instances, designations of one strand of a duplex and its complement (e.g., primer binding site and its complement) are arbitrary and could be reversed. For example, the complement of a primer binding site may itself be a primer binding site, and the complement of a loop segment is itself a loop segment.

[0042] The discrete components of the expression vectors are operably linked. "Operably linked" and "in operable linkage" means that the elements cooperatively function to achieve their intended task. For example, a promoter and a segment encoding an siRNA molecule are operably linked if the siRNA construct is transcribed when the promoter is activated. Similarly, a promoter and a transcription termination segment are operably linked, if

transcription of an RNA molecule initiated by the promoter is terminated when the RNA polymerase reaches the termination sequence.

[0043] The terms "sequence element" and "segment" may be used interchangeably, and both refer to a nucleic acid molecule of a certain sequence. For example, a tet operator sequence element may also be referred to as a tet operator segment. A "loop segment" can refer to the part of an siRNA molecule that is single stranded and connects the two annealed, complementary regions or alternatively can refer to the region of a DNA molecule that encodes this part of an siRNA molecule.

[0044] The term "binding site" refers to a section of DNA that is recognized and bound by another molecule in a sequence-specific fashion. For example, a primer binding site is a section of nucleic acid that is recognized by a primer nucleic acid molecule, preferably DNA, that anneals to the primer binding site. A restriction site is a section of DNA that is cleaved by a sequence-specific restriction endonuclease enzyme.

[0045] The term "regulatory element" means a DNA segment that is recognized in a sequence-specific fashion by a transcription factor protein. When the protein is bound to the regulatory element, if the regulatory element is located in or near a promoter, the promoter is either activated or repressed depending on the identity of the transcription factor.

[0046] Unless otherwise apparent from the context, all elements, steps or features of the invention can be used in any combination with other elements, steps or features.

DETAILED DESCRIPTION OF THE INVENTION

I. General

[0047] The invention provides methods of preparing a vector expressing an siRNA starting from a single template sequence encoding one strand of the siRNA. The strand of siRNA encoded by the template can be random. Methods of the invention are thus suitable for generating large libraries of random siRNAs in which one strand of each siRNA can be synthesized simultaneously in a solid phase synthesis procedure. Libraries of random siRNAs are useful tools for characterizing genes. By expressing random siRNAs in different cells, different genes are inactivated in the different cells. The resulting changes in phenotype can be correlated with the genes inactivated thereby characterizing the gene.

[0048] The invention also provides methods and constructs for expression of RNA molecules, including siRNA, from various promoters. A preferred expression system

comprises an inducible Pol III promoter in operable linkage with DNA encoding the RNA to be expressed and a Pol III termination site.

II. Methods of Producing siRNA Vectors

A. Strategy

[0049] The starting material for the methods is a DNA molecule having at least two components. One component is a segment encoding a strand of the siRNA to be expressed from the vector. The other component is a loop segment at the 3' end. The loop segment can self-anneal to form a hairpin loop. The DNA molecule containing these components is referred to in the claims as the first DNA molecule.

[0050] When self-annealed, the 3' end of the loop segment of the first DNA molecule can be extended by a polymerase using the remainder (i.e., non-self annealed portion) of the first DNA molecule as a template. Klenow or reverse transcriptase are particularly suitable polymerases for performing extension from a self-annealed template. Extension results in synthesis of a second DNA molecule fused to the 3' end of the first DNA molecule. The sequence of the second DNA molecule is complementary to the non-self annealed portion of the first DNA molecule which acted as a template in synthesis of the second DNA molecule. The combined second DNA molecule and first DNA molecule thus have at least the following components from 5' to 3': the segment encoding the strand of the siRNA, the loop segment, and the complement of the strand of the siRNA. The extension of a self-priming DNA molecule encoding one strand of an siRNA thus generates the second strand of the siRNA. The self-priming reaction leaves a duplex of second molecule bound to the complementary portion of the first molecule.

[0051] The duplex is then denatured to form a single-stranded molecule in which the second DNA molecule remains fused to the 3' end of the first DNA molecule. The denaturation can be effected by raising the temperature above the melting point or uses of denaturing chemical solvents. The single stranded molecule thus has a segment encoding a strand of an siRNA and its complement separated by the loop region.

[0052] The single-stranded DNA molecule serves as a template for synthesis of a second complementary strand forming a double stranded molecule. The double stranded molecule can be amplified by PCR. Optionally, second strand synthesis and PCR can be performed together. The resulting molecule is a double-stranded DNA fragment in which one strand

comprises the segments encoding a strand of an siRNA and its complement separated by the loop region, and the other strand is complementary to the first strand. The fragment or a subfragment thereof containing the siRNA strand and its complement are then subcloned into an expression vector under control of a promoter. Expression from the promoter produces a transcript comprising a segment having the siRNA strand sequence and its complement separated by a segment having the complement of sequence of the loop region. The transcript can self-anneal by pairing of the segment having the siRNA strand and its complement forming a double-stranded siRNA in which the two annealed strands are joined by the RNA segment having the loop sequence. Optionally, the loop region can be cleaved after expression by an endogenous or exogenously supplied RNA cleavage moiety (an endonuclease, e.g., Dicer, or a ribozyme).

[0053] Fig. 1 shows an exemplary scheme for performing the above method. The first DNA molecule (sscDNA) (1) has the following components from 5' to 3': a strand of a type II restriction endonuclease recognition site (RE1), a transcription termination segment (TTS/A5); the complement of a primer binding site (-CPBS) (overlapping the restriction endonuclease site and part of the transcription termination segment), a segment encoding a strand of an siRNA and a loop segment. The first DNA molecule is extended from the 3' end of the annealed loop segment using Klenow. Extension generates a second DNA molecule (C-sscDNA) fused to the 3' end of the first DNA molecule (sscDNA) (2). The components of the second DNA molecule are from 3' to 5': the complementary strand of the type II restriction endonuclease recognition site (CRE1), the complement of the transcription termination segment (CTTS/T5), the primer binding site (PBS) (overlapping the complement of the transcription termination segment (c-TTS/T5) and partially overlapping the complementary strand of the type II restriction endonuclease recognition site (CRE1), and the complement of the segment encoding the strand of siRNA (C-siRNA). The first and second DNA molecules form a duplex (dscDNA) whose two strands are joined by the loop segment .

[0054] The duplex (dscDNA) is then denatured to form a single stranded molecule (ssDNA) (3). The components of this single stranded molecule from 5' to 3' are: the restriction endonuclease cleavage site (-RE1), the transcription termination segment (TTS/A5), the complement of the primer binding site (CPBS), the segment encoding the a strand of the siRNA (-siRNA), the loop segment, the segment encoding the complement of the strand of the siRNA (C-siRNA), and the primer binding site (PBS), overlapping the complement of the transcription termination segment (CTTS/T5) and the complementary

strand of the type II restriction endonuclease recognition site (RE1). The duplex then undergoes second strand synthesis using a primer (4), followed by amplification using a pair of primers (4) and (7). The first primer (4) has a 5' tail comprising a strand of a recognition site for a type IIs restriction endonuclease (RE2).

[0055] Amplification generates a double stranded molecule with strands (5) and (6). Strand 6 is the same as single stranded molecule - (ssDNA) (3) other than strand 6 has an additional segment containing a complementary strand to the type IIs restriction endonuclease recognition site (CRE2) at its 3' end. Strand 5 is the strand complementary to the single-stranded DNA molecule (ssDNA) (3), having the following components from 5' to 3' order: the strand of the type IIs restriction endonuclease site (RE2), the strand of the type II restriction endonuclease site (RE1), the transcription termination segment (A5/TTS), a segment encoding the siRNA strand (siRNA), the complement of the loop segment of strand Y (C-Loop), a segment encoding the complement of the siRNA strand (C-siRNA), and a primer binding site (PBS) which is bound by a primer (7), containing the complement of the transcription termination segment (CTTS/T5)) and the complementary strand of the type II restriction endonuclease site (CRE1).

[0056] Amplified molecules containing strands 5 and 6 are then digested with enzymes RE1 and RE2, RE1 being a type II and RE2 a type IIs restriction enzyme. The digestion with RE2 is performed first. The enzyme cuts distal from its recognition site as shown by the arrow. The cleavage cuts both strands, thereby removing the RE1 recognition site and the transcription termination segment (A5) from the 5' end of strand 5. Cleavage by the RE1 enzyme then cuts both strands at the site shown by the vertical arrow at the right of the figure. The effect of the double digestion is to leave a double stranded molecule having ends generated by the RE2 cleavage at one end and by the RE1 cleavage at the other end. This double stranded molecule can be directionally cloned into an expression vector (8) having recognition sites for RE2 and RE1 downstream from a promoter.

[0057] Directional cloning results in transcription of the strand (6) of the DNA encoding siRNA. The transcript (9) comprises from 5' to 3' a strand of the siRNA, the complement of the loop segment, the complement to the strand of the siRNA and a short tail region (UU) formed by partial transcription of the termination segment.

B. Synthesis of the first DNA molecule

[0058] The first DNA molecule can be synthesized by solid state synthesis using a DNA synthesizer. Certain positions on the molecule can be randomized by including a pooled mixture of nucleotide for the corresponding coupling cycles. For complete randomization of a position all four natural nucleotides are present; two or three nucleotides can be present for partial randomization.

[0059] Alternatively, the first DNA molecule (ssDNA) can be synthesized by reverse transcription of an RNA precursor (ssRNA), which is in turn made by solid state synthesis. Fig. 2 shows an exemplary scheme for synthesizing the first DNA molecule (ssDNA) from an RNA precursor (ssRNA).

[0060] The RNA precursor (ssRNA) (10) has the following components from 5' to 3': a segment encoding the complement of the first DNA molecule loop sequence (C-Loop), a segment encoding the complementary strand of the siRNA strand sequence of the first DNA molecule (C-siRNA), and a primer binding site (PBS) overlapping the complement of a transcription termination segment (CTTS/T5) and a strand complementing the type II restriction endonuclease site of the first DNA molecule (CRE1). A DNA primer specific for the primer binding site of the RNA molecule (11) is annealed to the RNA molecule (ssRNA)(10) and extended via a reverse transcription reaction resulting in an RNA:DNA duplex (ssRNA:sscDNA) (13) of the RNA molecule (ssRNA) (10) and the complementary strand (12), the first DNA molecule (sscDNA)(1). The RNA molecule (ssRNA)(13) is digested from the duplex (ssRNA:sscDNA) via RNase H or an enzyme with similar activity, leaving the first DNA molecule (sscDNA) (1) ready for extension and amplification as shown in Fig. 1.

C. Characteristics of siRNAs

[0061] An siRNA has a double stranded region and optionally single stranded overhangs at the ends. The double stranded region is preferably between 10-30, 15-25, 17-23, 19-21, or 19-23 nucleotides in length. The double-stranded region is formed by pairing of segments (siRNA) and (C-siRNA) in Fig. 1. Single stranded regions, if present, are preferably 1-5 nucleotides in length. Single stranded overhangs can be introduced by partial transcription of the termination segment and/or asymmetric cleavage of the loop segment or its complement.

[0062] The siRNA can have defined sequences or can be fully or partially random. siRNAs having defined sequences have one strand (the antisense strand) that is fully or partly

complementary to a predetermined target gene of interest. Such siRNAs are used, for example, to suppress expression of the target gene. The sense strands of such siRNAs shows substantial sequence identity (e.g., at least 75, 80, 85, 95 and preferably 100%) to a segment of an RNA transcript sequence of a target gene. Any residues not complementary to the transcript sequence preferably occur in the 5' or 3' end of the siRNA.

[0063] siRNAs having fully random sequences are designed without regard to a particular target gene. In a strand of such an siRNA, each position (with the possible exception of single stranded overhangs) can be occupied by any of the four natural bases. Collections of random siRNAs are used, for example, in functional genomics to identify genes involved in detectable phenotypes of interest.

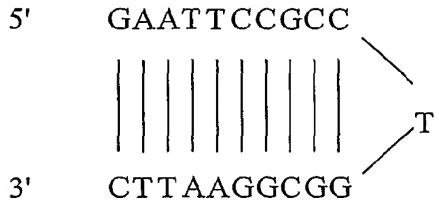
[0064] siRNAs having partially random sequences are designed such that at least one position can be occupied by at least two alternative bases. Partially random siRNAs are usually designed without a particular intended target gene sequence, but with a class of target genes, particular target organism or partial target gene sequence in mind. For example, to target mammalian cytochrome P450 genes, which have been shown to be GC rich in the third position of codons, increasing the proportion of G and/or C in the center segment of the first DNA molecule yields siRNAs with a higher likelihood of having affinity for mammalian cytochrome P450 genes. Conversely, the *Plasmodium falciparum* genome is particularly AT rich (~80%). Thus partially random siRNAs with a bias toward AT residues are more likely to recognize *Plasmodium* RNA transcripts than siRNAs having fully random sequences.

D. The Loop Segment

[0065] The first DNA molecule that forms the starting material for synthesis of an siRNA expression vector (see Fig. 1) contains a loop segment at the 3' end. The loop segment or its complement is carried through various intermediates to the final siRNA product. The loop segment comprises a short inverted repeat nucleotidesequence with one or more unpaired bases separating the inverted repeat regions. For example, in the sequence below GAATTCCGCC (SEQ ID NO:13) and GGCGGAATTC (SEQ ID NO:14) are inverted repeat regions and the central T is unpaired.

5' GAATTCCGCCTGGCGGAATTC 3' SEQ ID NO:12

The above loop segment (SEQ ID NO:12)can self-anneal to form:



[0066] Provided the loop segment is capable of self-annealing to a loop, its sequence is not critical. Optionally, the loop segment can be designed to be GC rich for greater stability. For example, the loop segment can have a GC content of at least 55%, 75%, or 90%, 95%. The loop segment should be sufficiently long to form a stable loop. Sometimes the loop segment is less than 10, 20 or 50 nucleotides in length. Formation of a loop can be confirmed by conformational gel analysis.

[0067] Optionally, the 3' loop segment is designed such that when annealed a cleavage site is present. The cleavage site can be that for a RNA endonuclease or a ribozyme. Optionally, the RNA endonuclease or ribozyme can be supplied exogenously by introducing a second vector encoding the endonuclease or ribozyme. The endonuclease or ribozyme is optionally under inducible control.

E. The primer binding site

[0068] In the exemplary scheme shown in Fig. 1, the first DNA molecule contains the complement of a primer binding site (CPBS). In the course of the scheme shown in Fig. 1 the complement of the primer binding site serves as a template for generation of a primer site used in second strand synthesis. The complement of the primer binding sites itself serves as a primer binding site in the amplification step of Fig. 1. The length and sequence of the priming site (and by implication its complement) are determined by conventional principles for primer selection. Optionally, the primer binding site overlaps fully or partially the Type II restriction endonuclease site and the transcription termination segment shown in the sscDNA molecule of Fig. 1.

F. Primers

[0069] The exemplary scheme shown in Fig. 1 employs two primers that bind to the primer binding site and its complement. Optionally, one of the primers (Primer A) contains a 5' tail that is not complementary to a template. The tail region contains one strand of a type IIs restriction endonuclease site. Type IIs restriction endonucleases have asymmetric

recognition sites that are distinct and usually upstream from their actual cleavage sites. Examples of type IIs restriction endonucleases are Fok I, Bsm BI, Bsa I, Bsm I, Sap I, Eco 57I, Bpm I, and Bsg I.. The type IIs endonuclease is chosen such that cleavage occurs between the complement of the transcription termination segment and siRNA segment of DNA molecule X, as shown in Fig. 1.

G. Transcription Termination Segment

[0070] In a double stranded DNA molecule, one strand of a transcription termination segment contains a run of about 5 adenine residues (A's) and the other strand a complementary run of about 5 thymidine residues (T's). The run of A's and T's is often immediately preceded (i.e., 5' on the strand containing the T's) with a GC rich region with palindromic symmetry. To effect termination of transcription of a double stranded DNA, the run of A's occurs on the transcribed strand, and the GC region is 3' to the A's on the transcribed strand. Transcription usually terminates within the run of A's leaving an RNA ending in one or more uridine (U) residues.

H. Subcloning of DNA encoding siRNA into an expression vector

[0071] The exemplary scheme shown in Fig. 1 shows one way of transferring a DNA molecule encoding both strands of an siRNA to an expression vector. In this scheme, a transcription termination segment is removed from one end of the molecule, and the resulting molecule is transferred into an expression vector by directional cloning such that the end of the DNA molecule lacking a transcription termination segment is closest to the promoter. Alternatively, the transcription termination segment can be left at the left end of molecule Z shown in Fig. 1 and the both ends of the molecule cut with the same restriction enzyme (RE1 in Fig. 1). The resultant fragment can insert in both orientations with respect to the promoter allowing transcription of both strands.

[0072] Both prokaryotic and eukaryotic expression vectors, such as mammalian, insect, fungal or viral, can be used. The particular expression vector used to transport the genetic information into the cell is selected with regard to the intended use of the siRNA, e.g., expression in plants, animals, bacteria, fungus, protozoa.

[0073] Elements typically included in the expression vectors include a promoter, an enhancer, an origin of replication, unique restriction sites in nonessential regions of the plasmid to allow insertion of recombinant sequences, and one or more genes encoding

selectable markers, e.g., antibiotic resistance, to permit selection of bacteria that harbor recombinant plasmids. A typical expression cassette within a vector contains a promoter operably linked to a downstream nucleic acid sequence encoding the siRNA, and signals required, e.g., for efficient polyadenylation of the transcript. Additional elements of the cassette may include, e.g., enhancers, and heterologous spliced intronic signals. In addition, some expression systems have markers for selection of stably transfected cell lines such as neomycin, thymidine kinase, hygromycin B phosophotransferase, and dihydrofolate reductase. Regulatory elements from eukaryotic viruses are often used in eukaryotic expression vectors, e.g., SV40 vectors, papilloma virus vectors, and vectors derived from Epstein-Barr virus. The elements that are typically included in the expression vectors also include a replicon that functions in E. coli, a gene encoding antibiotic resistance to permit selection of bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the plasmid to allow insertion of recombinant sequences.

[0074] General laboratory procedures for the cloning and expression of the siRNAs of the invention can be found, e.g., in current editions of Sambrook et al., Molecular Cloning A Laboratory Manual (2nd Ed.) Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989, or Current Protocols in Molecular Biology, Ausubel, ed. Greene Publishing and Wiley-Interscience, New York, New York (1987).

[0075] Suitable bacterial and eukaryotic promoters are described, e.g., in Sambrook et al., *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and Current Protocols in Molecular Biology (Ausubel et al., eds., 1994). The promoter used to direct expression of an siRNA nucleic acid depends upon the particular application. Either a constitutive or an inducible promoter is used, depending on the particular use of the clone encoding the siRNA. Exemplary eukaryotic promoters include the CaMV 35 S plant promoter, SV40 early promoter, SV40 late promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedron promoter, or other promoters shown effective for expression in eukaryotic cells. Pol III promoters including U6, H1, and 7SK are of particular use in the methods of the invention. Promotors and optionally other regulatory elements used can be selected to restrict expression to distinct cell types. For example thy-1, neural enolase, prion or PDGF promoters may be used to express siRNAs in cells of the central nervous system. The promoter can include elements that are responsive to transactivation, e.g., hypoxia response elements, GA14 response elements, lac repressor response element, and

small molecule control systems such as tet-regulated systems, the RU-486 system, ecdysone-regulated systems, and others (see, e.g., Gossen & Bujard, *Proc. Nat'l Acad. Sci. USA* 89:5547 (1992); Oligino et al., *Gene Ther.* 5:491-496 (1998); Wang et al., *Gene Ther.* 4:432-441 (1997); Neering et al., *Blood* 88:1147-1155 (1996); and Rendahl et al., *Nat. Biotechnol.* 16:757-761 (1998)).

[0076] Elements of the tet operon, which confers tetracycline resistance in *E. coli* and other gram negative bacteria, have been used to design systems for the regulation of gene expression in a wide variety of eukaryotic cells and organisms. These systems are based upon the strong affinity of the tet repressor protein (TetR) for the tet operator DNA sequences (tetO1, and tetO2) in the absence of tetracycline. This interaction inhibits the expression of the tetR and tetA genes, encoding the tet repressor and tet proteins respectively. Tet proteins are localized in the bacterial inner membrane and confer tetracycline resistance by catalyzing the efflux of the antibiotic as a complex with a divalent metal ion, such as Mg+2, in an energy dependent exchange for protons. Binding tetracycline induces a conformation change in the TetR protein, reducing its affinity for the tetO binding sites. The reduction of affinity results in the release of the TetR protein from the tet operator sequence, allowing expression of the tet and tet repressor proteins.

[0077] Tet-regulated systems can be used to control the expression of a variety of genetic constructs, including genes, gene fragments, or siRNAs, upon transfection into a wide variety of eukaryotic cells, including yeast, plant, amphibian, insect, and mammalian cells. As the tet elements are prokaryotic in origin, they have little effect on resident gene expression in target cells or tissue. Common tet-regulated systems include tet repression systems (Gossen and Bujard, PNAS USA 89:5547 (1992)), tet trans-activator systems (Gossen et al., *Science* 268:1766 (1995)), tet reversed activator systems (Gossen, *Biotechniques*, 19:213 (1995)), autoregulatory tet systems (Shockett et al., *Proc Nat Acad Sci* USA 92:6522 (1995)), and tet systems with reduced basal activity (Deutschle et al., *Mol. Cell. Bio.*, 15:1907 (1995)).

III. Preparation of siRNA Expression Libraries

[0078] The methods of the invention described in the preceding paragraphs are particularly suitable for the preparation random siRNA expression libraries. A fully or partly random library can be formed by pooling multiple nucleotides at one or more, or all of the positions of the siRNA strand within the first DNA molecule (as shown in e.g., Fig. 1). Typically all four nucleotides (A, C, G, and T) are included in the pools at each position but pools of fewer

nucleotides can also be used. If all four nucleotides are used at each of 19 positions in an siRNA, then $4^{19} = 2.7 \times 10^{11}$ different permutations result.

[0079] A population of nucleic acids encoding siRNA segments can be normalized to enrich the population for nucleic acids complementary to a nucleic acid library of interest. Enrichment is particularly useful for libraries of random siRNA segments in which most members are not complementary to a natural mRNA. Such an enrichment step reduces or eliminates siRNA segments lacking complementarity to any potential target mRNA. Screening of the enriched libraries increases the proportion of cells transformed with such libraries that undergo phenotypic changes due to siRNA suppression. Nucleic acids encoding siRNAs can be normalized at any stage of the method, but are preferably normalized before other extension or amplification steps are performed. Optionally, normalization can be performed on nucleic acid segments encoding siRNAs and a hairpin loop region added to the nucleic acid segments surviving enrichment, thus forming the first DNA molecules of the invention. The same normalization strategy can also be used in other methods of screening random nucleic acid sequences.

[0080] Normalization is performed, for example, by contacting nucleic acid molecules encoding siRNA strands with a nucleic acid library of interest under hybridization conditions. Hybridization generates double-stranded nucleic acid molecules in which nucleic acid segments encoding siRNAs are annealed to complementary sequences of the nucleic acid library of interest. Following hybridization, different methods can be performed to complete the normalization process.

[0081] In one method single-stranded nucleic acid molecules of the mixture (including nucleic acid encoding siRNAs that are not complementary to any sequences in the nucleic acid library of interest and the single-stranded overhangs of the double-stranded molecules) are subjected to enzyme-mediated degradation. The double-stranded molecules surviving degradation can then be denatured via heat or chemical-based methods. The siRNAs encoded by the resulting population of nucleic acid molecules are those capable of specifically binding to nucleotide sequences of the library of interest. A higher proportion of siRNAs prepared from the normalized nucleic acid molecules are more likely to recognize nucleotide sequences found in the library of interest than for siRNAs made from an unnormalized population of nucleic acid molecules.

[0082] The library used for normalization can be a DNA or RNA libraries, including mRNA and cDNA libraries. Libraries can be synthesized or obtained from commercial sources. Libraries of particular interest include species, or tissue, or developmental stage-specific libraries. Libraries from different sources can be combined to obtain a library containing substantially all expressed sequences from a given organism.

[0083] Figure 3 depicts an exemplary method for normalizing nucleic acids encoding siRNA molecules. The method is performed on a DNA segment encoding a random population of siRNA strands. The DNA segment is linked to other components as shown in Fig. 3. From 5' to 3', a biotin molecule (or other immobilization tag) is attached to a nuclease-resistant linker, the linker molecule is attached via a cleavable disulfide bond to a segment of nuclease-resistant DNA, which is attached to the 5' end of the DNA segment a random 21-mer siRNA strand (3A). The molecule is hybridized to an mRNA library of interest resulting in an RNA:DNA hybrid molecule (3B). The mixture is then subjected to single-stranded nucleic acid digestion, e.g., via S1 or mung bean nuclease, which degrades unannealed portions of the mRNA molecules and DNA molecules that have failed to hybridize to mRNA (3C). Optionally, the released DNA can be subject to further rounds of enrichment. The remaining duplexes are heat-denatured and the released DNA molecules of the invention are subjected to capture on a solid support such as streptavidin beads to remove them from the denatured complementary mRNA (3D). Following removal of mRNA, the hairpin adaptor (loop segment) can be ligated to the 3' end of the DNA molecule for subsequence formation of the complementary DNA sequence via Klenow enzyme activity. The biotin tag is removed by cleavage of the disulfide bond. Optionally, the cleavage is performed after other extension and amplification steps. Optionally, all steps can be performed in solution without use of an immobilizable tag or the disulfide bond.

[0084] Figures 11-12 depict an alternative method for normalizing nucleic acids encoding siRNA molecules. As in the previous normalization method, the nucleic acid library of interest can be from a variety of sources, such as mRNA or cDNA, in single-stranded or double-stranded form, and from one or more cell lines or tissues. A double-stranded library can used in double stranded form, or alternatively can be rendered to a single-stranded form by PCR, primer extension, melting, or other techniques. The nucleic acid library of interest can be used intact, or alternatively can be sheared physically or chemically, or can be digested with DNase I or a similar enzyme to yield smaller, more uniform length, fragments. In this alternative method, a nucleic acid library of interest is conjugated with an

immobilization tag that specifically binds to another molecule attached to a solid support. For example, a nucleic acid library of interest can be conjugated with biotin. When biotin is the conjugate, the nucleic acid library of interest is co-incubated with and thus immobilized onto streptavidin-coated particles such as magnetic beads or streptavidin-coated colloidal paramagnetic microbeads, as depicted in Figure 11. As shown in figure 12, the nucleic acid library of interest coated onto particles is then incubated with a randomized oligonucleotide library that is to be normalized. The buffer used for this binding step can be, for example, an array of salt and formamide concentrations, depending on the melting temperatures of the oligonucleotides of interest, to facilitate specific binding of the randomized oligonucleotide library to the immobilized nucleic acid library of interest and to minimize non-specific binding.

[0085] Preferably, the mixture of the randomized oligonucleotide library and the immobilized nucleic acid library of interest is denatured and allowed to renature. The denaturation/renaturation process can be performed, for example, by heating the mixture to a temperature sufficient to denature all binding between the nucleic acid library of interest and the immobilized nucleic acid library of interest. Typical temperatures for such denaturation are 50°C, 65°C, 75°C, 85°C, 95°C, and higher. Temperatures above 85°C are preferred when the immobilized nucleic acid library of interest and/or randomized oligonucleotide library is GC-rich, such as for example over 55% GC-rich. Temperatures above 85°C are also preferred when the immobilized nucleic acid library of interest and/or randomized oligonucleotide library contains hairpins or other secondary structures. Alternatively to heat, molecules such as formamide and urea can be used to denature the immobilized nucleic acid library of interest and the randomized oligonucleotide library. After denaturation, the immobilized nucleic acid library of interest and randomized oligonucleotide library are allowed to specifically bind. When temperature is the method used for denaturation, specific binding is achieved through cooling the mixture. Preferably the mixture is allowed to cool slowly, such as less than 1°C per minute until a desired temperature is reached. Slow cooling can be achieved by placing a tube containing the heated mixture in a thermocycler programmed to drop the temperature of the reaction at a pre-programmed rate. Alternatively, slow cooling can be achieved through placing a heated tube containing the mixture at room temperature. When chemical denaturation is used, specific binding is achieved through decreasing the concentration of the denaturing molecule. The concentration of the denaturing molecule can be decreased through methods such as dilution or dialysis.

[0086] After specific binding between the immobilized nucleic acid library of interest and the randomized oligonucleotide library is achieved, the mixture is typically washed with a buffer that does not disrupt the specific binding interactions but removes unbound randomized oligonucleotides. For example, when the nucleic acid library of interest is immobilized onto magnetic beads, a magnetic field is applied and the beads are immobilized within a column or tube while the unbound oligonucleotides washed away and removed by the buffer. Preferably, the bound beads are washed with buffers of progressively decreasing salt concentrations and/or increasing stringency to dissociate and wash away weakly bound or non-specifically bound randomized oligonucleotides. For example, the bound beads are washed with 300mM and then 100mM NaCl buffers. Weakly bound randomized oligonucleotides partially anneal to members of the nucleic acid library of interest but with less affinity than randomized oligonucleotides that have perfect complementarity with a member of the library. The specifically bound randomized oligonucleotides are then eluted with a low or no salt buffer, optionally containing formamide. The eluted and normalized randomized oligonucleotides are then used for subsequent production and cloning of a siRNA library, as described in other sections herein.

[0087] After synthesizing the first DNA molecule with pooled positions, all subsequent synthesis, and amplification steps are carried out in parallel on the many different random forms of the first DNA molecule. The end product is that DNA segments encoding different random siRNAs cloned into multiple copies of the same vector. The resulting library can be transformed into a population of cells. Usually only one vector goes into any one cell. Thus, different cells in the population receive vectors with different inserts and express different random siRNAs. The libraries of cells expressing different random siRNAs generated by this approach can have at least 10^6, 10^8 or 10^8 members.

[0088] Semi-random libraries can also be constructed that are biased toward targeting particular genes. For example, the random siRNA sequences can be synthesized using a solution of nucleotides that contains the same ratio of GC:AT nucleotides as a host cell that the library will be expressed in. Introducing this type of bias into the semi-random library increases the likelihood that a member of the library targets a transcript in the host cell. For example, the ratio of the frequency of GC:AT base pairs in the human genome is 0.41:0.59. An siRNA library constructed in which 20.5% of the nucleotides in the synthesis mixture are G, 20.5% are C, 29.5% are A, and 29.5% are T yields a larger percentage of members of the

library that by chance target a human gene than if all nucleotides in the synthesis mixture were present as 25% of the total.

IV. Methods of Delivering Nucleic Acids encoding siRNAs.

[0089] siRNAs can be delivered to cells (bacterial, fungal, insect, mammalian) or tissues using methods such as lipofection, microinjection, ballistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, agent-enhanced uptake of DNA, calcium phosphate transfection, polybrene, protoplast fusion, electroporation, plasmid vectors, viral vectors, both episomal and integrative. In general, it is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one nucleic acid sequence into the host cell capable of expressing the siRNA or choice.

[0090] Lipofection, cationic liposome-mediated transfection techniques, are described e.g., in U.S. Patent Nos. 5,049,386, US 4,946,787, and US 4,897,355. Additionally, lipofection reagents, including EasyFector, PrimeFector, GenePorter™, LipofectAMINE™ 2000, Lipofectamine PLUS™, CellFECTIN® are commercially available.

[0091] For biolistic techniques, DNA or RNA are precipitated to the surface of small (1-4 μm metal (e.g., gold or tungsten, depending upon the mode of acceleration used) particles. The particles are accelerated by one of several means, including movement through an electric field, explosive charge, or helium gas pulse, to penetrate tissues to a depth of several cell layers. Also known as "particle guns" or "gene guns", biolistic devices, such as the Helios Gene Gun from Bio-Rad, and the Accel Gene Delivery Device by Agacetus, Inc., are commercially available.

[0092] Nucleic acids encoding siRNA can also be delivered using viral vector delivery systems, including DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. The use of RNA or DNA viral based systems for the delivery of nucleic acids encoding randomized siRNAs take advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Conventional viral based systems for the delivery of siRNAs could include retroviral, lentiviral, adenoviral, adeno-associated, herpes-simplex virus and TMV-like viral vectors for gene transfer. Viral vectors are currently the most efficient and versatile method of gene transfer in target cells and tissues. Integration in the host genome is possible with the

retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

[0093] The tropism of retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retrovirus vectors that are able to transducer or infect non-dividing cells and typically produce high viral titers. Selection of a retrovirus gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of *cis*-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), murine stem cell virus (MSCV), gibbon ape leukemia virus (GaLV), simian immunodeficiency virus (SIV), human immuno-deficiency virus (HIV), and combinations thereof (see, e.g Buchscher et al., *J. Virol.* 66:2731-2739 (1992); Johann et al., *J. Virol.* 66:1635-1640 (1992); Sommerfelt et al., *Virol.* 176:58-59 (1990); Wilson et al., *J. Virol.* 63:2374-2378 (1989); Miller et al., *J. Virol.* 65:2220-2224 (1991); PCT/US94/05700).

[0094] In applications where transient expression of siRNAs is preferred, adenoviral based systems are typically used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors are also used to transducer cells with target nucleic acids (see, e.g., West et al., *Virology* 160:38-47 (1987); U.S. Patent No. 4,797,368; WO 93/24641; Kotin, *Human Gene Therapy* 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994)). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Patent No. 5,173,414; Tratschin et al., *Mol Cell. Biol.* 4:2072-2081 (1984); Hermonat & Muzyczka, *Proc. Nat'l Acad. Sci. USA* 81:6466-6470 (1984); and Samulski et al., *J. Virol.* 63:3822-3828 (1989).

[0095] Packaging cells are used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells , which package adenovirus, and ψ2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producer cell that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a

host, other viral sequences being replaced by an expression cassette for the protein to be expressed. The missing viral functions are supplied *in trans* by the packaging cell line. For example, AAV vectors typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely *rep* and *cap*, but lacking ITR sequences. The cell line is also infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by e.g, heat treatment to which adenovirus is more sensitive than AAV.

[0096] In many situations, it is desirable that the vector be delivered with a high degree of specificity to a particular cell type. A viral vector is typically modified to have specificity for a given cell type by expressing a ligand as a fusion protein with a viral coat protein on the viruses outer surface. The ligand is chosen to have affinity for a receptor known to be present on the cell type of interest. For example, Han et al., *Proc. Nat Acad. Sci. USA* 92:9747-9751 (1995), reported that Moloney murine leukemia virus can be modified to express human heregulin fused to gp70, and the recombinant virus infects certain human breast cancer cells expressing human epidermal growth factor receptor. This principle can be extended to other pairs of virus expressing a ligand fusion protein and target cell expressing a receptor. For example, filamentous phage can be engineered to display antibody fragments (e.g., FAB or Fv) having specific binding affinity for virtually any chosen cellular receptor. Although the above description applies primarily to viral vectors, the same principles can be applied to nonviral vectors. Such vectors can be engineered to contain specific uptake sequences thought to favor uptake by specific target cells.

V. Inducible Expression Systems

[0097] As described above, a preferred expression system expresses siRNA molecules from an inducible Pol III promoter. The same expression system can be used for any type of RNA including anti-sense RNA molecules. Use of a Pol III promoter is advantageous in achieving substantially uniform transcriptional termination. Pol III promoters can be made inducible as noted above, and described in further detail below by incorporation of at least one regulatory element into a Pol III promoter. A stimulus can act via the regulatory element to induce the Pol III promoter containing the regulatory element to initiate transcription of an

RNA molecule in operable linkage with the promoter. Transcription terminates in a substantially uniform manner at a transcription termination segment. Inducible Pol III promoters thus have the combined properties of substantially uniform termination and inducibility.

[0098] The increased homogeneity of RNA molecules expressed by Pol III promoters compared with e.g., Pol II promoters results in more uniform and reproducible phenotypic effects than if each RNA expressed from the construct were of variable length. For example, siRNAs expressed by Pol III promoters achieve more uniform gene silencing effects than when expressed by Pol II promoters due to their uniform length. Inducibility is also advantageous because it allows one to compare the time required for different changes of phenotype resulting from expression of a population of RNAs in different cells. For example, silencing of one gene by an siRNA construct can cause the desired phenotype to arise within 30 minutes of induction while silencing of a different gene by a different siRNA construct can cause the desired phenotype to arise only after several hours or days. Inducibility is also advantageous in analyzing phenotypes that would otherwise be lethal, since the level of induction can be modulated. For example, a tetracycline-inducible library of siRNA constructs can be induced with different levels of tetracycline to allow for identification of constructs that cause a phenotype at one level of induction but are lethal to the host cell at a higher level of induction.

[0099] The components of a vector containing an inducible Pol III promoter for expressing an siRNA are a Pol III promoter, a segment encoding an siRNA to be expressed, a transcription termination segment and one or more regulatory elements within the Pol III promoter. The segment encoding the RNA molecule to be expressed is positioned between the promoter and transcription termination segment, as depicted in Figure 9. The at least one regulatory element is positioned within the promoter as described in more detail below, and is depicted in Figure 8. The components are not usually linked to one another in nature. For example, an artificial siRNA construct may be placed downstream from an artificially constructed inducible Pol III promoter and upstream of a naturally occurring Pol III terminator segment.

[0100] Pol III promoters fall into three subtypes, generally known as type I, type II, and type III Pol III promoters. Type III Pol III promoters have three distinct sequence elements, known as a distal sequencing element (DSE), a proximal sequencing element (PSE), and a TATA box (see Paule et al., *Nucleic Acids Res.* Mar 15;28(6):1283-98 (2000)). For example,

the U6 Pol III promoter has these elements, positioned at approximately -230, -70, and -25, respectively, relative to the start of transcription. Other examples of type III Pol III promoters are H1, and 7SK. Types I and II Pol III promoters have different sets of sequence elements that recruit a different set of transcription factors (see Paule, Id.). Distinct sequence elements in promoters recruit transcriptional activators or repressors that affect the frequency with which the promoter is activated.

[0101] Here, a "Pol III promoter" means a promoter that recruits RNA Polymerase III to initiate transcription. The term includes naturally occurring type I, type II, or type III Pol III promoters, and variants thereof.

[0102] The distinct sequence elements of a Pol III promoter are interspersed by spacer DNA. The composition of spacer DNA is not critical. For example, there is significant sequence flexibility that can occur in the intervening nucleotides between the DSE, PSE, and TATA box elements of type III Pol III promoters. Regulatory elements can thus be inserted into spacer DNA without disrupting the capacity of a promoter to initiate transcription.

[0103] For example, a tet operator segment can be positioned between the PSE and the TATA box of a type III Pol III promoter, as depicted in Figure 8. The promoter is placed in operable linkage with a DNA segment encoding an mRNA molecule, which is in turn followed downstream by and in operable linkage with a termination segment. Binding of the tet repressor to the regulatory element between the PSE and the TATA box effectively blocks transcription initiation. The transcription factor(s) that bind to the PSE (such as SNAPc) and the TATA box (TATA binding protein) are unable to interact functionally when the tet repressor is bound to the regulatory element between the two sites. Addition of tetracycline to the culture media results in release of the tet repressor from the tet operator sequence. The tet repressor is released because of a dramatic reduction in its affinity for DNA upon binding tetracycline. Once the tet repressor has released from the tet operator sequence, the Pol III promoter becomes activated by recruitment of transcription factors to the PSE and TATA box. Two different tet operator sequences can be used, known as tet O1 and tet O2. Tet O2 is bound by the tet repressor with higher affinity than tet O1. The tet O1 sequence is ACTCTATCATTGATAGAGT (SEQ ID NO:15), whereas the tet O2 sequence is TCCCTATCAGTGATAGAGA (SEQ ID NO:16). As demonstrated in Example 2, use of tet O1 results in higher inducible expression but more "leakiness" compared to tet O2.

[0104] Other regulatory elements that can be placed in a promoter include an ecdysone response element, and a progesterone response element. When placed in a cell that expresses the ecdysone receptor (Ecr) and the product of the ultraspiracle gene (USP), an ecdysone response element recruits Ecr-USP heterodimers that activate transcription. Ecr-USP heterodimers only bind DNA and activate transcription when they are bound to the steroid hormone ecdysone. Ecdysone can therefore be used as a stimulus to activate a Pol III promoter containing an ecdysone response element. For reviews of inducible expression systems, see for example Ryding et al., *J. Endocrinol.* Oct;171(1):1-14 (2001); Harvey et al., *Curr. Opin. Chem. Biol.* Aug;2(4):512-8 (1998); and Albanese et al., *Semin. Cell Dev. Biol.* Apr;13(2):129-41 (2002). Alternatively, a progesterone response element can be inserted into a promoter and will impart inducibility to the promoter in the presence of RU-486 and the progesterone receptor. (see Wang et al., *Proc. Natl. Acad. Sci.* U S A. 1994 Aug 16;91(17):8180-4).

[0105] Transcription factors can also be fused to other domains such as transcriptional activation or repression domains. For example, the tet repressor gene can also be fused with a gene encoding a transcriptional activator. The resulting chimeric protein binds to tet operator sequences and activates transcription upon binding DNA (see Herr, *Cold Spring Harbor Symposia on Quantitative Biology* 63, 599-607 (1998); and Wang et al., *Proc. Natl. Acad. Sci. U S A.* 91(17):8180-4 (1994)).

[0106] For systems that involve the use of heterologous transcription factors such as tet repressor or Ecr and USP, the heterologous transcription factor should be present in the cell or in vitro transcription/translation mixture. For example, a cell line that is used for creating a library of constructs inducible by tetracycline can be first transformed with a constitutively expressing tet repressor construct, as described in Example 2. Alternatively, the tet repressor cDNA can be included in the Pol III construct used for expressing siRNA.

[0107] Any DNA segment can be placed into the construct for expression by the promoter. Examples include genes encoding proteins, RNA genes such as snRNA and tRNA, as well as DNA segments or artificial genes, such as those encoding antisense and siRNA molecules. The discrete length of the expressed RNA molecules results from the utilization of Pol III transcriptional terminators. RNA Pol III terminates transcription after the second or third thymine upon encountering a stretch of four or more thymine nucleotides surrounded by a GC-rich nucleotide context. These sequences (such as GGAGATGAGATCTTTTTTCTCGAG (SEQ ID NO:17)), known as terminators, terminator

segments, termination segments, or termination sequences, are placed downstream of the gene or construct to be expressed. The GC-rich sequence is frequently palindromic, but this is not a prerequisite for effective transcriptional termination. The result of Pol III recognizing the termination segment and releasing the nascent RNA is that each RNA molecule expressed by the construct has a discrete size. For example, a 100 base pair siRNA molecule expressed using a Pol III promoter and Pol III termination segment is generated almost exclusively as 100 and 101 nucleotide RNA molecules if the second thymine in the terminator sequence is nucleotide number 100. A variety of transcriptional terminator sequences can be placed in the construct, such as the U6 terminator. For descriptions of other Pol III terminator sequences, see Gunnery et al., *J Mol Biol.* Feb 26;286(3):745-57 (1999) and Chu et al., *J Mol Biol.* Feb 26;286(3):745-57 (1999).

[0108] By contrast, a population of RNA molecules transcribed by a Pol II promoter is more heterogeneous in length because Pol II recognizes polyadenylation signals as a means of terminating transcription. Once Pol II recognizes a polyadenylation signal, it terminates transcription and allows polyA polymerase to add adenine ribonucleotides to the 3' end of the transcript. Anywhere from 10 to over 200 adenine ribonucleotides are added to each Pol II-generated transcript. Because polyA polymerase adds adenine ribonucleotides without any complementary template to read from, and because the cofactor polyA-binding protein (PABP) variably influences the ultimate length of the transcript, polyA polymerase adds a different number of adenine ribonucleotides to transcripts expressed from the same physical piece of DNA.

VI. Direct Insertion of siRNA molecules into cells

[0109] Although DNA encoding siRNAs are usually transformed directly into cells, it is possible to express DNA encoding siRNAs in vitro, isolate the siRNA expression products, and then insert the siRNA molecules into cells. Insertion of siRNA molecules into cells can be accomplished using techniques such as lipofection, electroporation,a nd microinjection. siRNAs can be transformed using many of the techniques applicable to DNA encoding an siRNA. The siRNA molecules inserted into cells using these methods can be synthesized in vitro, for example, using a Pol III promoter-Pol III terminator expression vector.

VII. Applications

[0110] siRNA expression vectors produced by the methods of the invention have a wide variety of uses including research, diagnostic and therapeutic applications.

A. Characterization of gene function

[0111] siRNAs produced by methods of the invention can be used to characterize gene function by determining a property resulting from suppression or silencing of a gene. The property can be a phenotype detectable by visualization of the cell or a molecular change detectable by molecular analysis of a cell extract. Phenotypes that can be correlated with gene silencing or suppression include changes in level of protein(s) or mRNAs, proliferation rates of transformed cells vs. untransformed controls, immortality, changes in anchorage dependence, changes in growth factor requirements, foci formation, growth in soft agar, apoptosis assays, e.g., DNA laddering and cell death, expression of genes involved in apoptosis; signal transduction assays, e.g., changes in intracellular calcium, cAMP, cGMP, IP3, changes in hormone or neurotransmitter release; receptor assays, e.g., estrogen receptor and cell growth; growth factor assays, e.g., EPO, hypoxia and erythrocyte colony forming unit assays; enzyme production assays, e.g., FAD -2 induced oil desaturation and respiration assays, pathogen resistance assays, e.g., insect, bacterial, parasite, and viral resistance assays; chemical production assays, e.g., penicillin production, transcription assays, e.g., reporter gene assays; and protein production assays, e.g., VEGF ELISAs, and changes in gas requirements and temperature requirements. For example, the genes responsible for a desired property can be determined by a subtractive hybridization technique. Cells transformed with an siRNA library are screened for the desired property. cDNA libraries are made from those transformed cells having the desired property These libraries are used as drivers to subtract complementary DNA from cDNA libraries prepared from untransformed cells. The remaining DNA is enriched for cDNAs whose lack of expression is associated with the desired property. These cDNAs can be sequenced and analyzed by standard molecular biology techniques.

[0112] As an alternative or additional approach, one can isolate vector encoding siRNA from clonal isolates of cells having the desired property. The siRNA encoding region of such vectors is sequenced. The sequence is then compared with genomic sequences in a database to identify genes that might be regulated by the siRNA. The lack of expression of these genes is then associated with the desired property. The sequence of the siRNA can also be used to design probes or primers to screen a cDNA or genomic library for complementary DNA segments. Suppression of expression of these DNA segments is also associated with the desired property.

[0113] As another approach one can compare patterns of mRNA expression in transformed cells having a desired property with control untransformed cells of the same type. DNA probe arrays and associated materials for this purposes are commercially available from Affymetrix among others. This approach detects both mRNAs whose expression is suppressed by an siRNA and mRNAs whose expression levels changes (either up or down) responsive to suppression of other mRNAs. The differentially expressed mRNAs identified by this analysis are associated with the desired property. In the case of mRNAs expressed at lower levels in transformed cells with the desired property than control cells, then lack of expression of the mRNA correlates with the desired property.

[0114] The procedure used to identify cells having a desired property depends on the nature of the desired property. If the property is presence or absence of a cell surface antigen, the property can be detected uses conventional immunological binding assays, such as ELISA. If the property is cell proliferation rate, cells with the desired property can be identified using tritiated thymidine or tritiated 5-bromo-deoxyuridine incorporation. Other properties, such as resistance to a drug, or lack of dependence on a nutritional supplement can be selected by propagating transformed cells in the presence of the drug or absence of the nutritional supplement. Other properties, such as cell morphology or anchorage dependence can be detected by visual or miscroscopic observation.

[0115] The same type of experiment can be performed using a single siRNA known to be complementary to a gene of unknown function with a view to characterizing the gene. In this situation one transforms a population of cells with vectors encoding the same siRNA and then studies clonal isolates of transformed cells for altered properties of types such as discussed above.

[0116] Use of an inducible Pol III promoter is advantageous for gene identification assays. For example, a population of cells can be transformed with a library of random siRNA sequences cloned into inducible Pol III expression constructs. The integration of the random siRNA library construction and the inducible Pol III promoter construction creates random siRNA libraries that are simultaneously induced in individually isolated cell transformants. For example, a random siRNA library cloned into inducible Pol III expression vectors can be transformed into a population of cells. A selectable marker in the vector allows colonies that have retained the construct to be picked and arrayed into multi-well plates. Each well contains a clonal population of cells containing an expression vector that can inducibly express a randomly generated siRNA sequence. An exogenous stimulus is added to the cells

to stimulate transcription of the siRNA vectors, such as tetracycline. A desired phenotype is screened or selected for by comparing each independent transformant colony to each other and to control cells.

B. Therapeutic Applications

[0117] siRNAs produced using the methods of the invention also have therapeutic applications, particularly in the area of gene therapy. An expression vector encoding an siRNA can be used to prevent the translation of genes causing undesirable or deadly conditions such as viral infections, malignancies, or dominant inherited diseases. In such methods, the siRNA typically as a defined sequence chosen to be complementary to a particular target gene to be suppressed.

[0118] For example, bone marrow cells from an HIV-infected individual can be collected and transformed with one or more siRNA expression vectors, thereby producing one or more HIV virus-specific siRNAs. The transformed bone marrow cells, now "educated" to destroy HIV mRNA transcripts, are then reintroduced to the patient's bone marrow where they can replicate, forming a cell population capable of protecting against HIV. Other viruses that can be attacked using this approach include hepatitis, herpes and human papilloma viruses.

[0119] In other methods, siRNAs are designed to suppress expression of mutant oncogenes without affecting expression of a wildtype allele. Transformation with expression vectors encoding such siRNAs is useful for suppressing cancers caused by expression of the mutant oncogene.

[0120] Expression of siRNAs from an inducible Pol III promoter is advantageous for clinical applications. Inducible Pol III promoters can be cloned into gene therapy vectors. Such vectors produce siRNA molecules only when induced and only of a specifically designed size. This reduces a source of clinical variability based on transcriptional differences between cells of the same type in different patients and different cell types in the same patient. It also reduces heterogeneity of siRNAs expressed from the same vector.

[0121] siRNAs can be designed to suppression expression of dominant inherited diseases. Some examples of dominant inherited diseases are achondroplasia, Alagille syndrome, Alzheimer disease (AD), amyotrophic lateral sclerosis (ALS), BRCA1 and BRCA2 hereditary breast/ovarian cancer, familial idiopathic basal ganglia calcification (FIBGC), branchiootorenal (BOR) syndrome, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), Charcot-Marie-tooth neuropathies

types 1 and 2(CMT types 1 and 2), congenital contractural arachnodactyly (CCA), craniosynostosis syndromes (FGFR-Related), DRPLA, GTP cyclohydrolase 1 (GTPCH1)-deficient Dopa-responsive dystonia (DRD), early-onset primary dystonia (DYT1), facioscapulohumeral muscular dystrophy (FSHD), familial adenomatous polyposis (FAP), frontotemporal dementia with Parkinsonism-17 (FTDP-17), glucose transporter type 1 deficiency syndrome (Glut1-DS), Greig cephalopolysyndactyly syndrome (GCPS), hereditary hemorrhagic telangiectasia, hereditary sensory neuropathy type I (HSN1), hereditary spastic paraplegia (HSP), Huntington disease (HD), hypochondroplasia, hypokalemic periodic paralysis (HypPP), incontinentia pigmenti (IP), Li-Fraumeni syndrome (LFS), Marfan syndrome, multiple endocrine neoplasia type 2 (MEN 2), hereditary multiple exostoses (HME), myotonic dystrophy (DM), nemaline myopathy (NM), Neurofibromatosis Types 1 and 2 (NF1 and NF2), nevoid basal cell carcinoma syndrome (NBCCS), autosomal dominant nocturnal frontal lobe epilepsy (ADNFLE), DFNA3 (Connexin 26), Noonan syndrome (NS), oculopharyngeal muscular dystrophy (OPMD), *PTEN* hamartoma tumor syndromes (PHTS), Pallister-Hall syndrome, Peutz-Jeghers syndrome (PJS), autosomal dominant polycystic kidney disease (ADPKD), retinitis pigmentosa (RP), retinoblastoma (RB), Rubinstein-Taybi syndrome (RSTS), pinocerebellar ataxia types 1, 2, 3, 6, 7, 8, and 10 (SCA1, SCA2, SCA3, SCA6, SCA7, SCA8, and SCA10), Stickler syndrome, transthyretin amyloidosis, tuberous sclerosis complex (TSC), Von Hippel-Lindau syndrome (VHL syndrome), Waardenburg syndrome type 1 (WS1), and Williams syndrome (WS).

C. Other applications (non-siRNA)

[0122] In other methods, genetic libraries are synthesized to induce expression of other molecules such as a diverse set of mRNA molecules. Use of inducible Pol III promoters and Pol III transcriptional terminators allows for each mRNA to have an identical 3' untranslated region and eliminates differential phenotypic effects based on transcript stability arising from different lengths of 3' untranslated regions. Antisense molecules can be expressed in a similar manner to siRNAs. Once colonies transformed with antisense RNAs have been identified as having a desired phenotype, the vector encoding the gene or artificial construct that causes the desired phenotype can be isolated and sequenced as described above.

[0123] All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

[0124] Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to one of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

Example 1

[0125] The inducibility of tet operator sequences was tested by insertion of tet O1 and tet O2 elements into a Pol II (CMV) promoter for the expression of a green fluorescent protein (GFP) cDNA sequence. The design of the construct is depicted in Figure 4. Two tet O2 segments were placed immediately downstream of the TATA box in the CMV promoter. When tet repressor is bound, the promoter is not active. Upon the addition of Tetracycline to HeLa cells containing this construct, the tet repressor releases from the tet O2 segments and allows transcription of the GFP gene.

[0126] The results of the tet-GFP experiment are shown in Figure 5. 5A shows fluorescence with (center panel) or without (bottom panel) the addition of Tetracycline. 5B shows confocal microscopy shots of the same fields of cells as cells shown in 5A.

[0127] Quantitative PCR was used to determine the relative expression of GFP in HeLa cells with and without tetracycline added. Two sets of cells were transfected with the CMV-GFP construct. One set also expresses the tet repressor (HeLa::TR) while the other set does not (HeLa), as depicted in Figure 6. Cells cultured in the absence of tetracycline (-T) or with tetracycline (+T) are indicated. C3 denotes those cells transfected with a control vector from which GFP is expressed under the control of a pol II promoter (CMV). TOeGFP denotes those cells transfected with a GFP-expressing vector where GFP is expressed under the control of a tet-repressible pol II promoter (CMV-TO). The data indicate that only in cells expressing the tet-repressor is GFP expression inducible.

Example 2

[0128] An inducible Pol III - siRNA expression vector system was constructed to demonstrate inducible expression of RNA molecules from Pol III promoters with a downstream Pol III terminator. The construct was designed as a portable cassette which can be cloned, for example, with restriction sites or by Gateway recombination when flanked by the appropriate sites, into any vector of interest. For example, the cassette can be moved into

the appropriate vector for either transient transfection, or retrovirus, lentivirus, or adenovirus infection. Due to its portable nature, it is not limited to these vector platforms.

[0129] The cassette was constructed in pBluescript II SK by PCR cloning with primers specific for the U6, H1, and 7SK Pol III promoters. In the case of the U6 cassette used in these experiments, PCR primers of SEQ ID NO: 1 and SEQ ID NO: 2 were used to amplify the human U6 promoter from a preparation of HeLa cell genomic DNA. The PCR product was gel purified, digested with BamHI and PSTI, and cloned into pBluescript II SK cut with BamHI and PSTI. The sequence of the cloned U6 promoter PCR product is SEQ ID NO: 3. H1, and 7SK Pol III promoters were cloned in a similar manner.

[0130] Subsequently, the resulting pBluescript II SK-U6 vector was cut with PSTI and XhoI. The oligonucleotides of SEQ ID NOs: 4 and 5 were annealed, digested with PSTI and XhoI, and the products gel purified and ligated with PSTI/XhoI-cut pBluescript II SK-U6 vector to yield the constitutive U6 cassette, depicted in Figure 7. The cassette can be subcloned using BamH1 and XhoI into any desired vector. Digestion with the appropriate restriction enzyme (eg PstI or EcoRV and BseR1) allows directional cloning of an adapted hairpin oligo to allow directional cloning of a 3'-TT overhang. For example, for a PstI and BseR1 digest, and to target the sequence AA-N(19), the synthetic hairpin sequences can be:

5'-GGAT-N(19)-TT- CAAGAGA (loop)-N'(19)-TT-3' as the sense oligo (SEQ ID NO:18)

5'-N(19)-TCTCTTGAA-N'(19)-ATCCTGCA-3' as the antisense oligo (SEQ ID NO:19)

[0131] The two oligos are annealed and cloned into the PstI and BseR1 digested U6 vector. Following introduction of the cassette into cells either alone or using an appropriate vector transcription of the following hairpin with a TT-3' overhang results:

5'-N(19)-TT-CAAGAGA-N'(19)-TT-3' (SEQ ID NO:20)

[0132] The constitutive U6 cassette was rendered inducible by performing site directed mutagenesis between the TATA box and the PSE to introduce tetracycline operator sites (tetO1 or tetO2), as depicted in Figures 8 and 9. Figure 8 depicts the placement of the tet O1 and tet O2 segments between the PSE and TATA box. A desired siRNA sequence can be cloned into the restriction sites such that the U6 promoter is upstream of the siRNA sequence and the U6 terminator is downstream of the siRNA sequence. For example, a randomized siRNA library can be cloned into the U6 siRNA cassette. Site directed mutagenesis was also used to create other convenient restriction sites for incorporation of appropriately adapted

siRNA expression sequences. The sequence of the U6 tet O1 expression cassette is SEQ ID NO: 6. The sequence of the U6 tet O2 expression cassette is SEQ ID NO: 7. Figure 9 is a schematic depiction of a tet-inducible U6-siRNA expression vector.

[0133] The tet-inducible U6-siRNA expression vector was tested in HeLa cells. Expression of an RNA segment cloned into the vector was measured by quantitative PCR. Quantitative PCR protocols are known in the art (for example, see Mol Biochem Parasitol. 2003 May;128(2):205-15.) The RNA segment is referred to as K2X, see SEQ ID NO: 8. In Figure 10, TO1K2X represents those cells transfected with a vector carrying the tet-inducible U6-siRNA expression vector with a tet O1 segment in the promoter (described above). TO2K2X represents those cells transfected with a vector carrying the Tet-inducible U6-siRNA expression vector with a Tet O2 segment in the promoter (also described above). The primers used in the quantitative PCR are SEQ ID NOs: 9 and 10, corresponding to sense and antisense primers, respectively, that amplify K2X. The quantitative PCR of the K2X sequence was performed using the Taqman® method, with SEQ ID NO: 11 as the dual labeled probe. Methods of quantitative PCR using the Taqman® methods are well known. For examples, see Ryu et al., *Microbiol Immunol.*;47(10):693-9 (2003) and Perelygina et al., *J Virol Methods.* ;109(2):245-51 (2003).

[0134] The data indicate that the K2X segment is inducibly expressed in cells carrying the tet O1 and tet O2 inducible U6-siRNA expression vector containing the K2X segment. Lanes 1-6 of Figure 10 correspond to HeLa cells expressing a tet repressor cDNA, while lanes 7-12 correspond to HeLa cells not transfected with a tet repressor cDNA vector. Lanes 1-2 and 7-8 are controls not containing the K2X segment. Lanes 3-4 demonstrate that the K2X segment, cloned into a tet O1-modulated vector (SEQ ID NO: 6) is expressed at a 2.7-fold higher level upon addition of tetracycline to the cells than cells in the absence of tetracycline. Lanes 5-6 demonstrate that the K2X segment, cloned into a tet O2-modulated vector (SEQ ID NO: 7) is expressed at a 4.1-fold higher level upon addition of tetracycline to the cells than cells in the absence of tetracycline. The data indicated that tet O2 is able to more tightly repress the expression of the K2X segment than tet O1.

Example 3

Generation and immobilization of a library for normalization of a random oligonucleotide pool:

[0135] A circular single-stranded library of known content was created by digestion of a plasmid of known sequence. This was accomplished by digestion of the circular single-

stranded plasmid with 50 mU of DNase I for 30 minutes at 37°C. The library was digested such that the majority of DNA fragments were below 500 bases in length as determined by gel electrophoresis on a 2% agarose gel. The resulting DNA was fractionated to remove the majority of fragments below 40 nucleotides in length. This linearized single-stranded DNA library of known sequence was end-labeled with biotin using terminal transferase (Roche). Excess unincorporated biotin was removed and the biotinylated single-stranded DNA library was immobilized onto streptavidin coated paramagnetic microbeads. Non-biotinylated single-stranded DNA was removed by suspending the beads in a microcolumn exposed to a magnetic field and washing away unbound material with 10mM Tris, 1mM EDTA, pH 7.4.

Oligonucleotide binding specificity to immobilized Library:

[0136] Two pools of oligonucleotides were synthesized. The first pool was designed to hybridize to the single-stranded DNA library. This pool contained 21mer and 35mer oligonucleotides that are 100% complementary to sequences within the library. The second pool was designed to not hybridize to the single-stranded DNA library. This pool contained 30mer and 25mer oligonucleotides that do not have complementarity to any sequence contained within the library. A 41mer oligonucleotide was also designed that contained a stretch of 21 nucleotides at the 3'-end that have 100% identity to a sequence within the library; however, the 20 nucleotides at the 5'-end do not have sequence identity to the library. The library on paramagnetic microbeads was immobilized on a microcolumn exposed to a magnetic field and was washed three times with one volume (100uL) of 10mM Tris, 1mM EDTA, pH 7.4. A 12% polyacrylamide/urea gel was used to visualize the oligonucleotides eluted from the microcolumn. The results of the experiment are shown in figure 13. Lane A is a control showing the migration pattern and quantity of the five oligonucleotides described above. A mixture of the five oligonucleotides described above was resuspended in 10mM Tris, 1mM EDTA, pH 7.4 and mixed with the library immobilized on paramagnetic beads. The temperature was raised to 85°C and cooled (0.5 degrees per second) to 21 degrees. The column was then washed with three washes (100uL each) of 10mM Tris, 1mM EDTA, pH 7.4. The column was eluted with 100uL of elution buffer (10% formamide, 10mM Tris, 1mM EDTA, pH 7.4). The eluent was collected and run in Lane B. The absence of oligonucleotides in Lane B demonstrates that under these buffer conditions, no oligonucleotides bind to the library. In an additional part of the experiment, the microbead-bound library was applied to a microcolumn in a magnetic field and the same mixture of five

oligonucleotides was resuspended in a binding buffer containing 300mM NaCl, 10mM Tris, 1mM EDTA, pH 7.4. The temperature was raised to 85°C and cooled (0.5 degrees per second) to 21 degrees. The column was then washed three times with 300mM NaCl, 10mM Tris, 1mM EDTA, pH 7.4, followed by three washes with a higher stringency buffer containing 100mM NaCl, 10mM Tris, 1mM EDTA, pH 7.4 to remove non-specifically bound oligonucleotides. As an alternative to stepped stringency buffers, a gradient of decreasing stringency buffer can be applied to the immobilized nucleic acid library. The column was eluted with elution buffer (10% formamide, 10mM Tris, 1mM EDTA, pH 7.4) and heated to 70°C. The eluent was collected and run on Lane C, which contains the three specific 21, 35, and 41mer sequences. The migration of these oligonucleotides is not directly parallel with the markers in lane A because the gel exhibited a "smile" pattern due to heat/salt conditions. Subsequently, the column was further eluted with three more volumes of elution buffer heated to 70°C. The eluent was collected and run on Lane D. These data demonstrate (i) no oligonucleotides bind to the library in the absence of salt (Lane B); (ii) only oligonucleotides specific for sequences within the library are still bound following stringency washes (Lane C); and (iii) all of the bound oligonucleotides are eluted in the first eluent fraction. These data demonstrate the ability to specifically bind and elute sequences found within an immobilized nucleic acid library of interest.

Sequences

SEQ ID NO: 1 (ck2089)
5'GAATTCGGATCCAAGGTCGGGCAGGAAGAGGGCCTA-3'

SEQ ID NO: 2 (ck2090)
5'AGGCCTAAGCTTCTGCAGTTTCCACAAGATATATAAAG-3'

SEQ ID NO: 3.:

GGATCCAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATA
TTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTA
AACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGG

TAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAAC
TTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACTGCAG

SEQ ID NO: 4ck2085-
5'GATCCTGCAGGATATCGAATTCAAGCTTGAGGAGATGAGATCTTTTTTCTCGAG
ACTACC-3'

SEQ ID NO: 5 ck2086-
5'GGTAGTCTCGAGAAAAAAGATCTCATCTCCTCAAGCTTGAATTCGATATCCTGC
AGGATC-3'

SEQ ID NO: 6:

CCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATA
ATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTA
GAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGCT
ATCATATGCTTACCGTAACTTGAAAGTACTCTATCATTGATAGAGTATATATCTTG
TGGAAAGGACTGCAGGATATCGAATTCAAGCTTGAGGAGATGAGATCTTTTTTCT
CGAGGGGGGGCCCGGT

SEQ ID NO: 7:

CCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATA
ATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTA
GAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGCT
ATCATATGCTTACCGTAACTTGAAAGTCCCTATCAGTGATAGAGATATATATCTT
GTGGAAAGGACTGCAGGATATCGAATTCAAGCTTGAGGAGATGAGATCTTTTTTC
TCGAGGGGGGGCCCGGT

SEQ ID NO: 8:

CCATGGATCGATACAGCGAGGTGGATCCGCTTGTCACTGCGCTGGATCGC
CGCGTAGATGCCCTCTTGGATGTCTCCGAAATGTGGCTTCAAGCTT

SEQ ID NO: 9:

CACATTTCGGAGACATCCAAGAG

SEQ ID NO: 10:

GAGGTGGATCCGCTTGTCA

SEQ ID NO: 11:

TCTACGCGGCGATCCAGCGC

SEQ ID NO: 12:

GAATTCCGCCTGGCGGAATTC

(SEQ ID NO:13)

GAATTCCGCC

(SEQ ID NO:14)

GGCGGAATTC

(SEQ ID NO:15)

ACTCTATCATTGATAGAGT

(SEQ ID NO:16)

TCCCTATCAGTGATAGAGA

(SEQ ID NO:17)

GGAGATGAGATCTTTTTTCTCGAG

(SEQ ID NO:18)

GGATNNNNNNNNNNNNNNNNNNTTCAAGAGAN'N'N'N'N'N'N'N'N'N'N'N'N'N'N'N'N'N'N'N'TT

(SEQ ID NO:19)

NNNNNNNNNNNNNNNNNNNTCTCTTGAAN'N'N'N'N'N'N'N'N'N'N'N'N'N'N'N'N'N'N'N'ATCCTGCA

(SEQ ID NO:20)

NNNNNNNNNNNNNNNNNNNTTCAAGAGAN'N'N'N'N'N'N'N'N'N'N'N'N'N'N'N'N'N'N'N'TT

WHAT IS CLAIMED IS:

1. A method of preparing a vector for expression of an siRNA, comprising:

(a) providing a first DNA molecule comprising a segment encoding a strand of an siRNA and a loop segment at the 3' end,

(b) extending the 3' end of the loop with a polymerase to form a second DNA molecule duplexed with the first DNA molecule, the second DNA molecule encoding the complementary strand of the siRNA,

(c) denaturing the duplexed first and second DNA molecules to form a single-stranded DNA molecule comprising the second DNA molecule fused to the 3' end of the first DNA molecule;

(d) synthesizing a second strand complementary to the single-stranded DNA molecule forming a double stranded DNA molecule; and

(e) cloning at least a fragment of the double stranded molecule into a vector in operable linkage with a promoter whereby transcription of the vector forms a transcript comprising a segment having the siRNA strand sequence, the loop segment or its complement, and the complement of the siRNA strand, and the transcript self-anneals to form an siRNA.

2. The method of claim 1, wherein the providing step is performed by synthesizing the first DNA molecule using solid-phase synthesis.

3. The method of claim 1, wherein the first DNA molecule further comprises the complement of a primer binding site at the 5' end, the second DNA molecule comprises the primer binding site at its 3' end, and step (d) is performed using a primer binding to the primer binding site.

4. The method of claim 3, wherein the first DNA molecule further comprises a transcription termination segment optionally within the complement of the primer binding site, and the second DNA molecule further comprises the complement of the transcription termination segment.

5. The method of claim 3, wherein the first DNA molecule further comprises a strand of a restriction enzyme recognition site optionally within the complement of the primer binding site.

6. The method of claim 1, further comprising amplifying the double stranded DNA molecule of step (d).

7. The method of claim 4, wherein the primer binding to the primer binding site on the second DNA molecule further comprises a 5' segment comprising a strand of a type IIs restriction endonuclease recognition site whereby digestion of the double stranded DNA molecule with the type IIs restriction endonuclease removes the complement of the transcription termination segment from the second DNA molecule.

8. The method of claim 1, wherein the siRNA strand has a random sequence.

9. The method of claim 1, wherein the method is performed on a plurality of first DNA molecules encoding different siRNA strand sequences, each of the different sequences being a random sequence, and the method forms a plurality of vectors encoding different siRNAs.

10. The method of claim 9, further comprising transforming the plurality of vectors into a population of cells and identifying which cells develop a change in a property responsive to suppression of expression of a gene by an siRNA.

11. The method of claim 3 wherein the providing step comprises:

(i) providing an RNA molecule comprising a segment encoding the complementary strand of the siRNA strand sequence flanked at its 5' end by an RNA segment having the complement of the loop sequence, and at its 3' end by an RNA segment having the complement of the primer binding site sequence;

(ii) extending a primer bound to the RNA segment having the complement of the primer binding site sequence with reverse polymerase to generate the first DNA molecule complexed with the RNA molecule; and

(iii) digesting the RNA molecule to leave the first DNA molecule free of the RNA molecule.

12. The method of claim1, wherein the first DNA molecule of step (a) is formed by ligation of the loop segment to the 3' end of the segment encoding a strand of an siRNA.

13. The method of claim 9, wherein the siRNA strands are a normalized population selected from a random population by hybridization to a nucleic acid library.

14. The method of claim 9, further comprising contacting the DNA segments encoding the siRNA strands with a nucleic acid library under hybridization conditions, and selecting the DNA segments that hybridize to the nucleic acid library.

15. The method of claim 14, wherein members of the library are conjugated to an immobilizable tag and the selecting comprises:

(i) immobilizing the nucleic acid library to a solid support;

(ii) contacting the DNA segments with the immobilized nucleic acid library under hybridization conditions;

(iii) washing the immobilized nucleic acid library to remove unhybridized or weakly hybridized DNA segments; and

(iv) eluting the hybridized DNA segments.

16. The method of claim 15, wherein the immobilizable tag is biotin..

17. The method of claim 15, wherein the washing comprises adding progressively lower salt concentrations of wash buffer to the immobilized nucleic acid library.

18. The method of claim 17, wherein the salt is NaCl.

19. The method of claim 14, further comprising linking the DNA segments that hybridize to the nucleic acid library to the loop segment to provide the plurality of first DNA molecules.

20. The method of claim 13, wherein the nucleic acid library is an mRNA library.

21. The method of claim 13, wherein the nucleic acid library is a cDNA library.

22. A method of enriching a random nucleic acid library, comprising:

(a) hybridizing the random nucleic acid library to an mRNA or cDNA population; and

(b) isolating members of the random nucleic acid library that bind to the population.

23. The method of claim 22, wherein the random nucleic acid library is a library of DNA segments encoding random RNA sequences.

24. The method of claim 23, wherein the library of DNA segments is linked to an immobilizable tag.

25. The method of claim 24, wherein the immobilizable tag is biotin.

26. The method of claim 24, wherein the DNA segments are linked to the immobilizable tag via a nuclease-resistant DNA segment, a disulphide bond, and a nuclease resistant linker.

27. The method of claim 26, further comprising contacting hybrids of the DNA segments and mRNA or cDNA population with a single-stranded nuclease to digest unhybridized DNA segments and unhybridized members of the mRNA or cDNA population.

28. The method of claim 27, further comprising subjecting the hybrids to denaturing conditions to separate the DNA segments from the members of the mRNA or cDNA population to which the DNA segments hybridized.

29. The method of claim 28, further comprising immobilizing the DNA segments via the immobilizable tag.

30. The method of claim 29, further comprising cleaving the disulfide bond to separate the DNA segments from the immobilizable tag.

31. The method of claim 23, wherein each member of the mRNA or cDNA population is linked to an immobilizable tag.

32. The method of claim 30, wherein the immobilizable tag is biotin.

33. A DNA construct comprising:

(a) a Pol III promoter;

(b) a segment encoding an RNA molecule;

(c) a transcriptional termination segment; and

(d) at least one regulatory element;

wherein the promoter, the segment encoding an RNA molecule, the termination segment and the at least one regulatory element are in operable linkage such that a stimulus can act via the at least one regulatory element to induce the promoter to initiate transcription of the segment encoding an RNA molecule, transcription being terminated by the transcriptional termination segment.

34. The construct according to claim 33, wherein the RNA molecule is an siRNA.

35. The construct according to claim 33, wherein the promoter contains a tet operator segment.

36. The construct according to claim 33, wherein the promoter is a type III Pol III promoter.

37. The construct according to claim 36, wherein a tet operator segment is located in the promoter.

38. The construct according to claim 37, wherein the tet operator segment is tet O1.

39. The construct according to claim 37, wherein the tet operator segment is tet O2.

40. The construct according to claim 33, wherein the transcriptional termination segment is a U6 Pol III terminator.

41. A library of RNA expression vectors, wherein each vector contains:

(a) a Pol III promoter;

(b) a segment encoding an RNA molecule, the RNA molecule sequence differing between vectors; and

(c) a transcriptional termination segment; and

(d) at least one regulatory element;

wherein the promoter, the segment encoding an RNA molecule, the termination segment and the at least one regulatory element are in operable linkage such that a stimulus can act via the at least one regulatory element to induce the promoter to initiate

transcription of the segment encoding an RNA molecule, transcription being terminated by the transcriptional termination segment.

42. The library according to claim 41, wherein each promoter contains a tet operator segment.

43. The library according to claim 41, wherein each promoter is the U6 promoter.

44. A method of expressing an RNA molecule comprising:

(a) providing a cell containing a DNA construct comprising:

(i) a Pol III promoter;

(ii) a segment encoding an RNA molecule;

(iii) a transcriptional termination segment; and

(iv) at least one regulatory element;

wherein the promoter, the segment encoding an RNA molecule, the termination segment and the at least one regulatory element are in operable linkage such that a stimulus can act via the at least one regulatory element to induce the promoter to initiate transcription of the segment encoding an RNA molecule, transcription being terminated by the transcriptional termination segment; and

(b) providing the stimulus, which acts via the regulatory element to induce the Pol III promoter.

45. The method according to claim 44, wherein the promoter contains a tet operator segment, the cell expresses the tet repressor, and the stimulus is provided by removing tetracycline from the culture media.

46. A method of identifying a gene comprising:

(a) synthesizing a library of random RNA segments;

(b) cloning the segments into a population of DNA constructs containing:

(i) a Pol III promoter;

(ii) a segment encoding an RNA molecule;

(iii) a transcriptional termination segment; and

(iv) at least one regulatory element;

wherein the promoter, the segment encoding an RNA molecule, the termination segment and the at least one regulatory element are in operable linkage such that

a stimulus can act via the at least one regulatory element to induce the promoter to initiate transcription of the segment encoding an RNA molecule, transcription being terminated by the transcriptional termination segment;

(c) transforming a population of cells with the population of DNA constructs containing the random RNA segments;

(d) providing a stimulus that acts via the at least one regulatory element to induce the promoter; and

(e) screening or selecting for a desired phenotype.

47. The method of claim 46, wherein the promoter contains a tet operator segment.

48. The method of claim 46, wherein the promoter is the U6 promoter.

49. The method of claim 46, wherein the promoter is the U6 promoter and a tet operator segment is located between the proximal sequencing element and the TATA box.

50. The method of claim 46, further comprising the steps of:

(a) recovering a DNA construct from a cell which exhibits the desired phenotype; and

(b) sequencing the RNA segment contained in the construct.

51. The method of claim 46, wherein the RNA segments are siRNA.

Figure 1

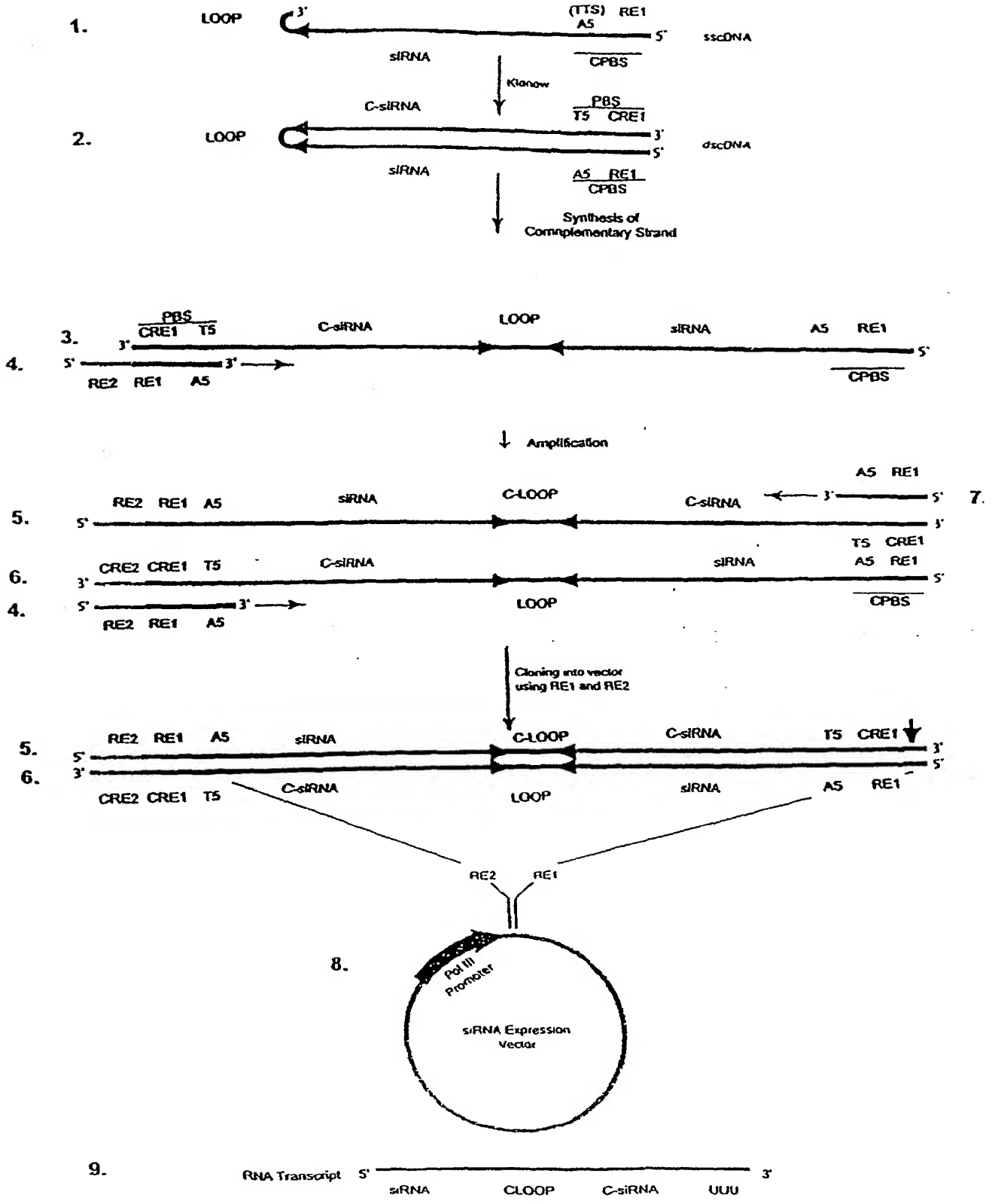

1.
LOOP
3'
(TTS) RE1
A5
5'
ssDNA
siRNA
CPBS
Klenow
C-siRNA
PBS
T5 CRE1
2.
LOOP
3'
5'
dsDNA
siRNA
A5 RE1
CPBS
Synthesis of
Complementary Strand
3.
PBS
CRE1 T5
3'
C-siRNA
LOOP
siRNA
A5 RE1
5'
4.
5'
3'
RE2 RE1 A5
CPBS
Amplification
5.
RE2 RE1 A5
siRNA
C-LOOP
C-siRNA
A5 RE1
7.
3'
5'
TS CRE1
6.
CRE2 CRE1 T5
C-siRNA
siRNA
A5 RE1
4.
LOOP
CPBS
Cloning into vector
using RE1 and RE2
5.
RE2 RE1 A5
siRNA
C-LOOP
C-siRNA
T5 CRE1
6.
CRE2 CRE1 T5
C-siRNA
LOOP
siRNA
A5 RE1
RE2
RE1
8.
Pol III
Promoter
siRNA Expression
Vector
9.
RNA Transcript
5'
3'
siRNA
CLOOP
C-siRNA
UUU

Figure 2

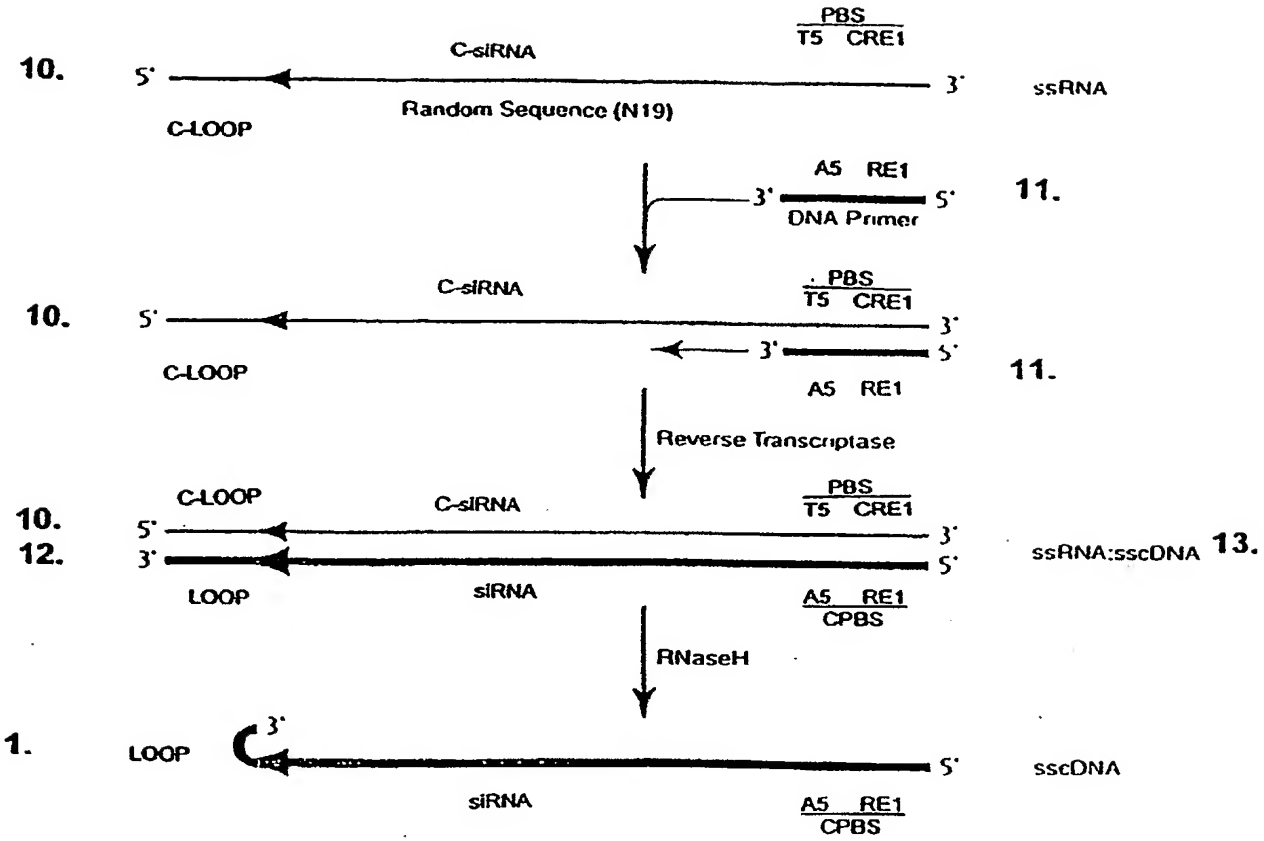

10.
C-siRNA
PBS
T5 CRE1
5'
3'
ssRNA
Random Sequence (N19)
C-LOOP
A5 RE1
11.
3'
5'
DNA Primer
C-siRNA
PBS
T5 CRE1
10.
5'
3'
3'
5'
C-LOOP
11.
A5 RE1
Reverse Transcriptase
C-LOOP
C-siRNA
PBS
T5 CRE1
10.
5'
3'
12.
3'
5'
ssRNA:sscDNA
13.
LOOP
siRNA
A5 RE1
CPBS
RNaseH
1.
LOOP
3'
5'
sscDNA
siRNA
A5 RE1
CPBS

Figure 3

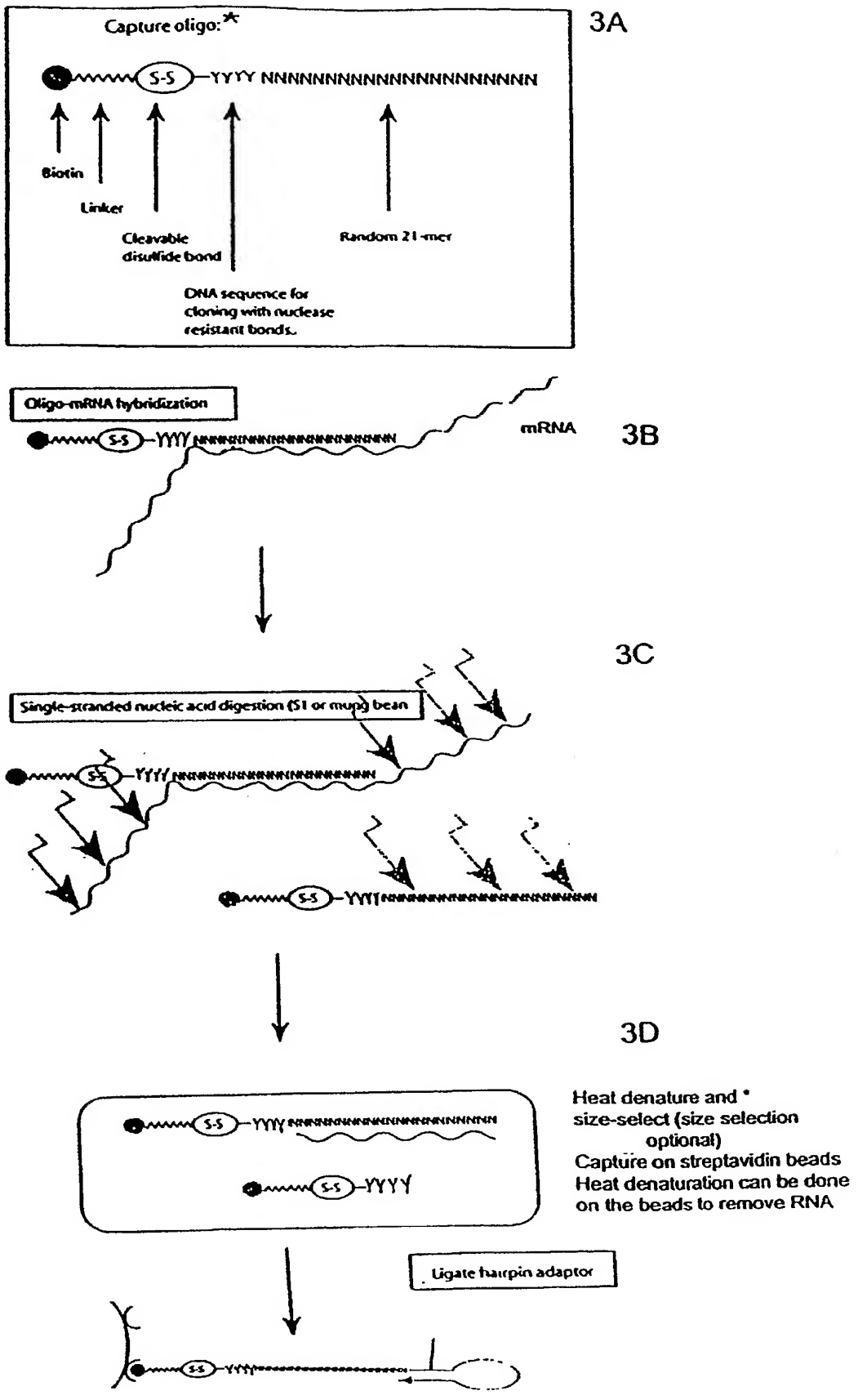
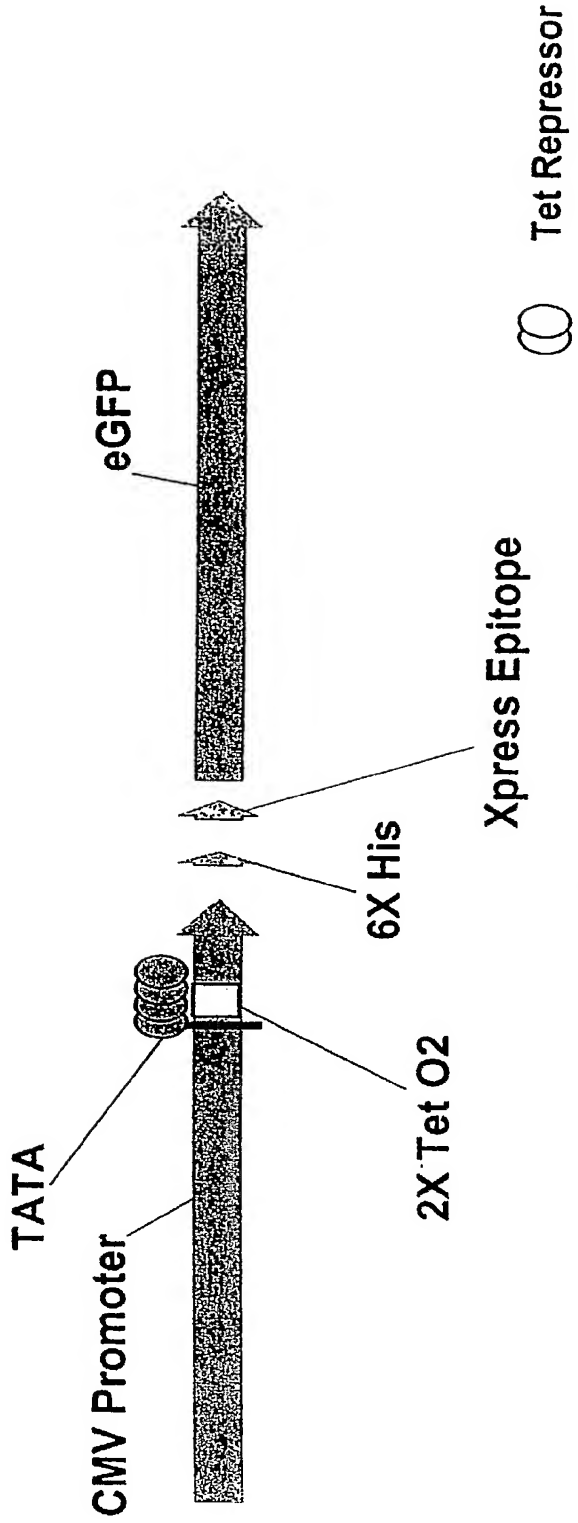

Capture oligo:*
S-S
YYYY NNNNNNNNNNNNNNNNNNNN
Biotin
Linker
Cleavable disulfide bond
DNA sequence for cloning with nuclease resistant bonds.
Random 21-mer
3A
Oligo-mRNA hybridization
mRNA
3B
3C
Single-stranded nucleic acid digestion (S1 or mung bean
3D
Heat denature and *
size-select (size selection optional)
Capture on streptavidin beads
Heat denaturation can be done on the beads to remove RNA
Ligate hairpin adaptor

Figure 4


pcDNA4/TO-eGFP
eGFP
Tet Repressor
Xpress Epitope
6X His
2X Tet O2
TATA
CMV Promoter

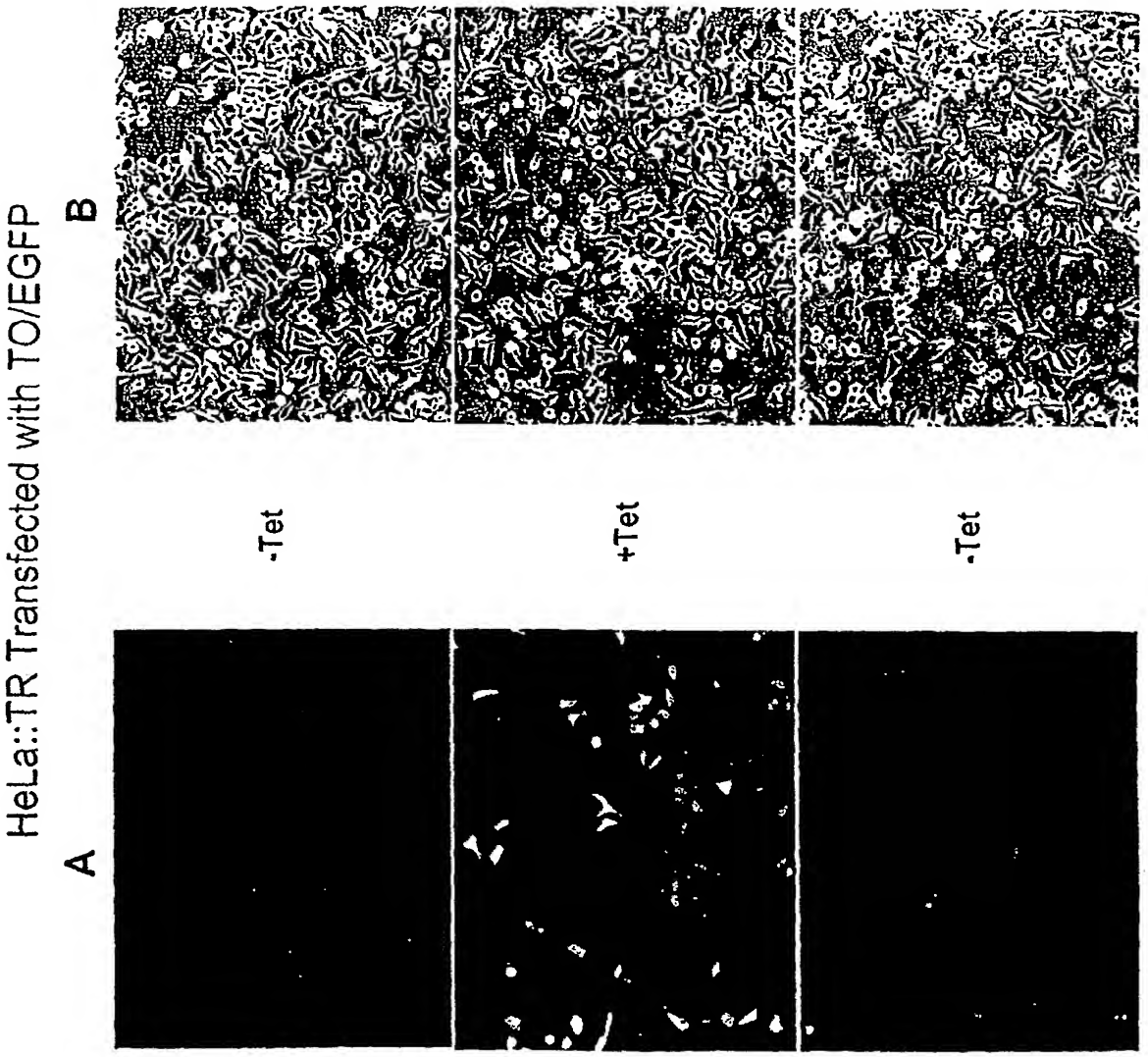

HeLa::TR Transfected with TO/EGFP
A
B
-Tet
+Tet
-Tet
TO/Null
TO/EGFP
TO/EGFP

Figure 5

Figure 6

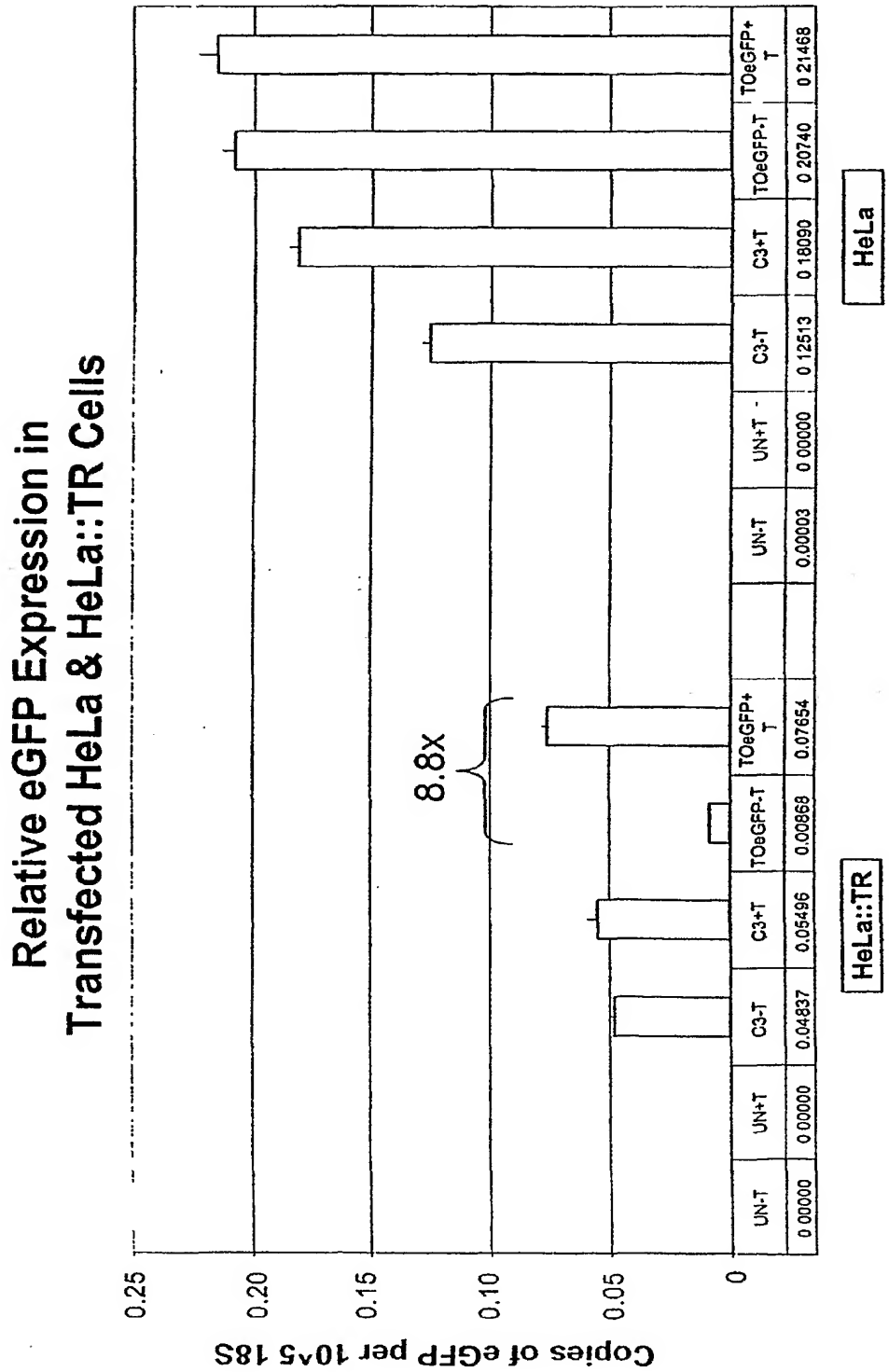

Relative eGFP Expression in Transfected HeLa & HeLa::TR Cells
Copies of eGFP per 10^5 18S
0.25
0.20
0.15
0.10
0.05
0
8.8x
UN-T
UN+T
C3-T
C3+T
TOeGFP-T
TOeGFP+T
UN-T
UN+T -
C3-T
C3+T
TOeGFP-T
TOeGFP+T
0.00000
0.00000
0.04837
0.05496
0.00868
0.07654
0.00003
0.00000
0.12513
0.18090
0.20740
0.21468
HeLa::TR
HeLa

Figure 7

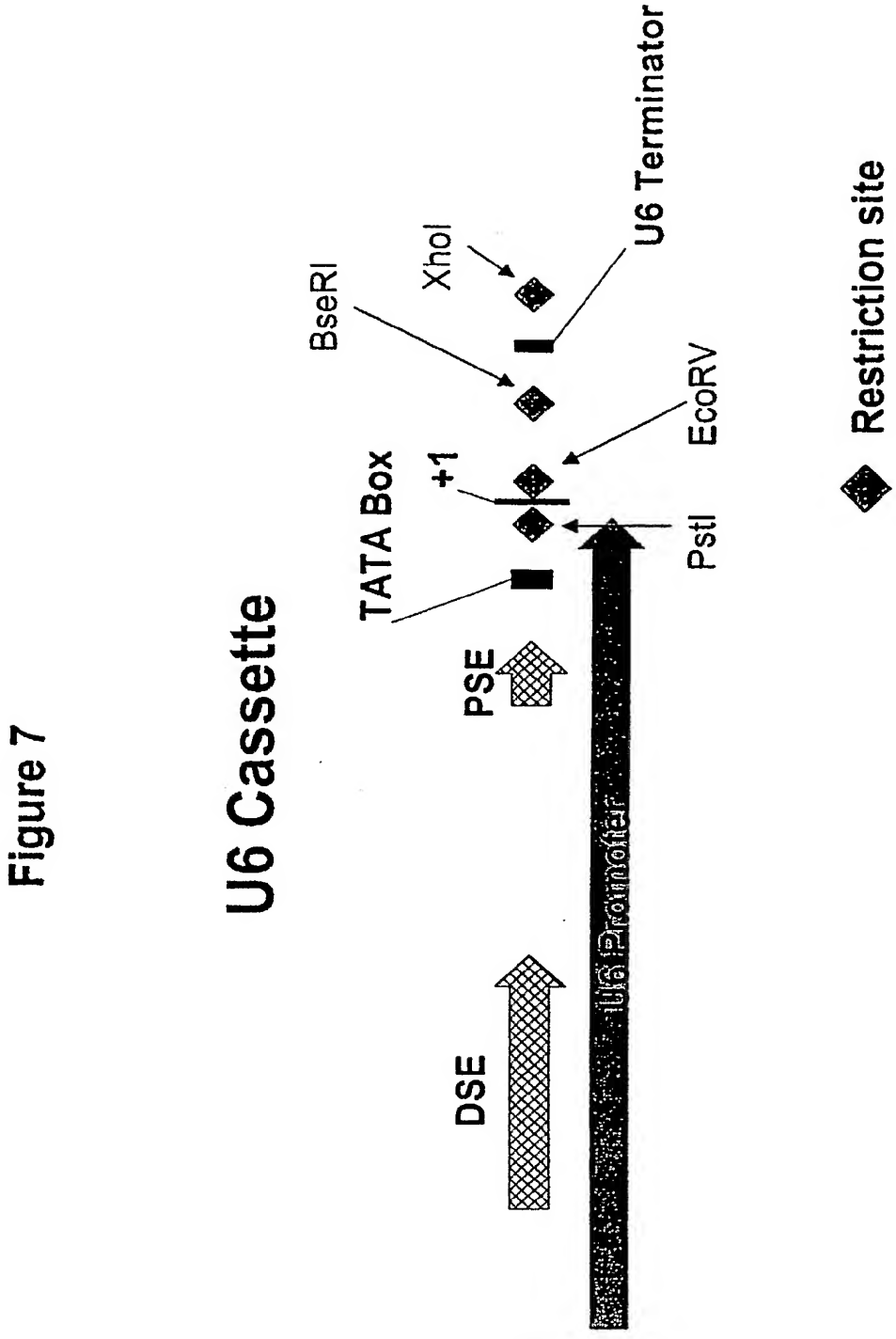

U6 Cassette
DSE
PSE
TATA Box
+1
BseRI
XhoI
U6 Terminator
EcoRV
PstI
U6 Promoter
Restriction site

Figure 8

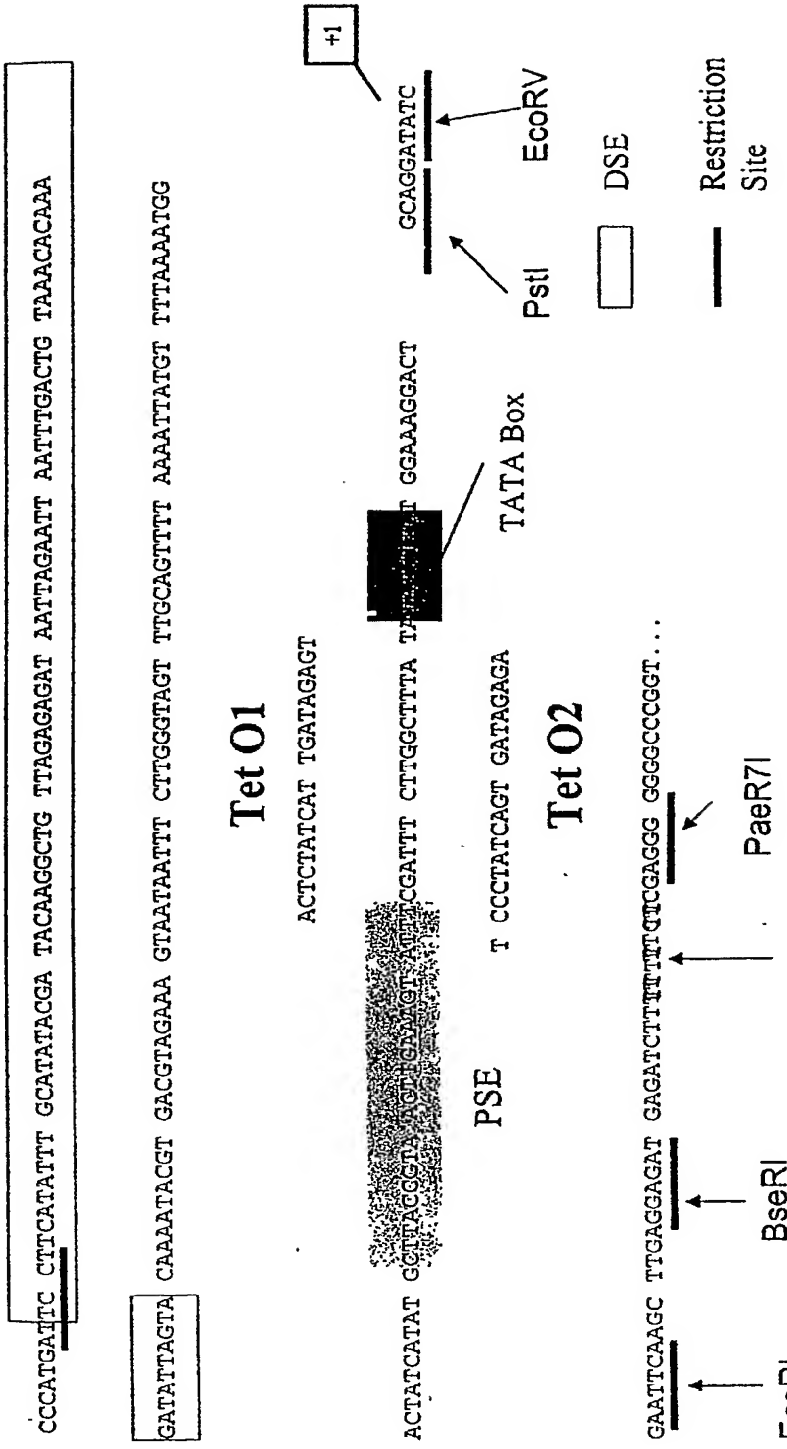

Inducible Pol III Promoter Construction
-237...
CCCATGATTC CTTCATATTT GCATATACGA TACAAGGCTG TTAGAGAGAT AATTAGAATT AATTTGACTG TAAACACAAA
-157...
GATATTAGTA
CAAAATACGT GACGTAGAAA GTAATAATTT CTTGGGTAGT TTGCAGTTTT AAAATTATGT TTTAAAATGG
Tet O1
ACTCTATCAT TGATAGAGT
-77...
ACTATCATAT
GGTTACGGTA AGTTCAATCT AGTTACGATT CTTGGCTTTA TATAAGGAAT GGAAAGGACT
PSE
T CCCTATCAGT GATAGAGA
TATA Box
Tet O2
+1
GCAGGATATC
PstI
EcoRV
DSE
Restriction Site
4...
GAATTCAAGC TTGAGGAGAT GAGATCTTTTTTTTCTCGAGGG GGGGCCCGGT...
EcoRI
BseRI
PaeR7I
Pol III Terminator

Figure 9

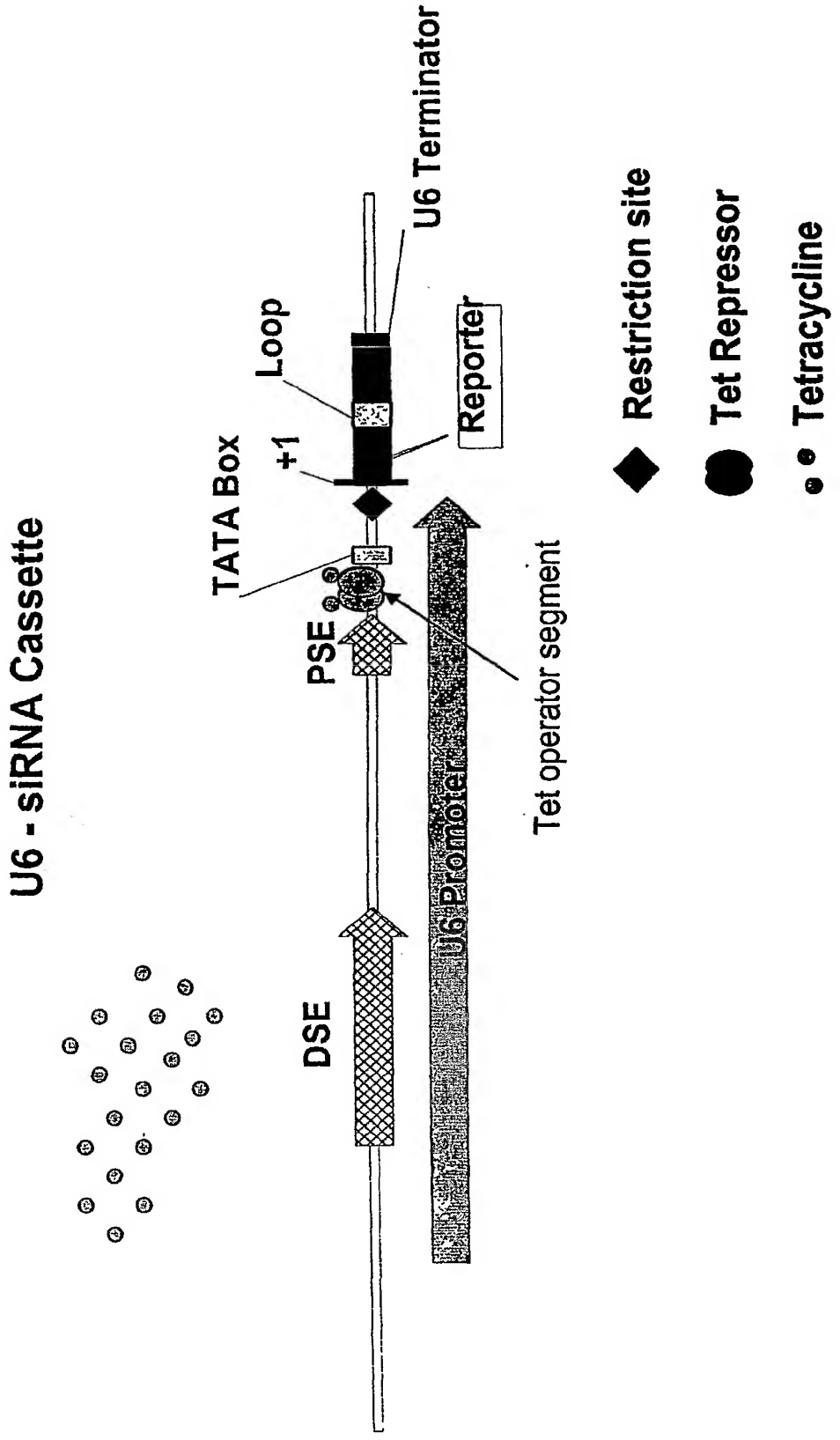

U6 - siRNA Cassette
DSE
PSE
TATA Box
+1
Loop
U6 Terminator
Reporter
U6 Promoter
Tet operator segment
Restriction site
Tet Repressor
Tetracycline

Figure 10

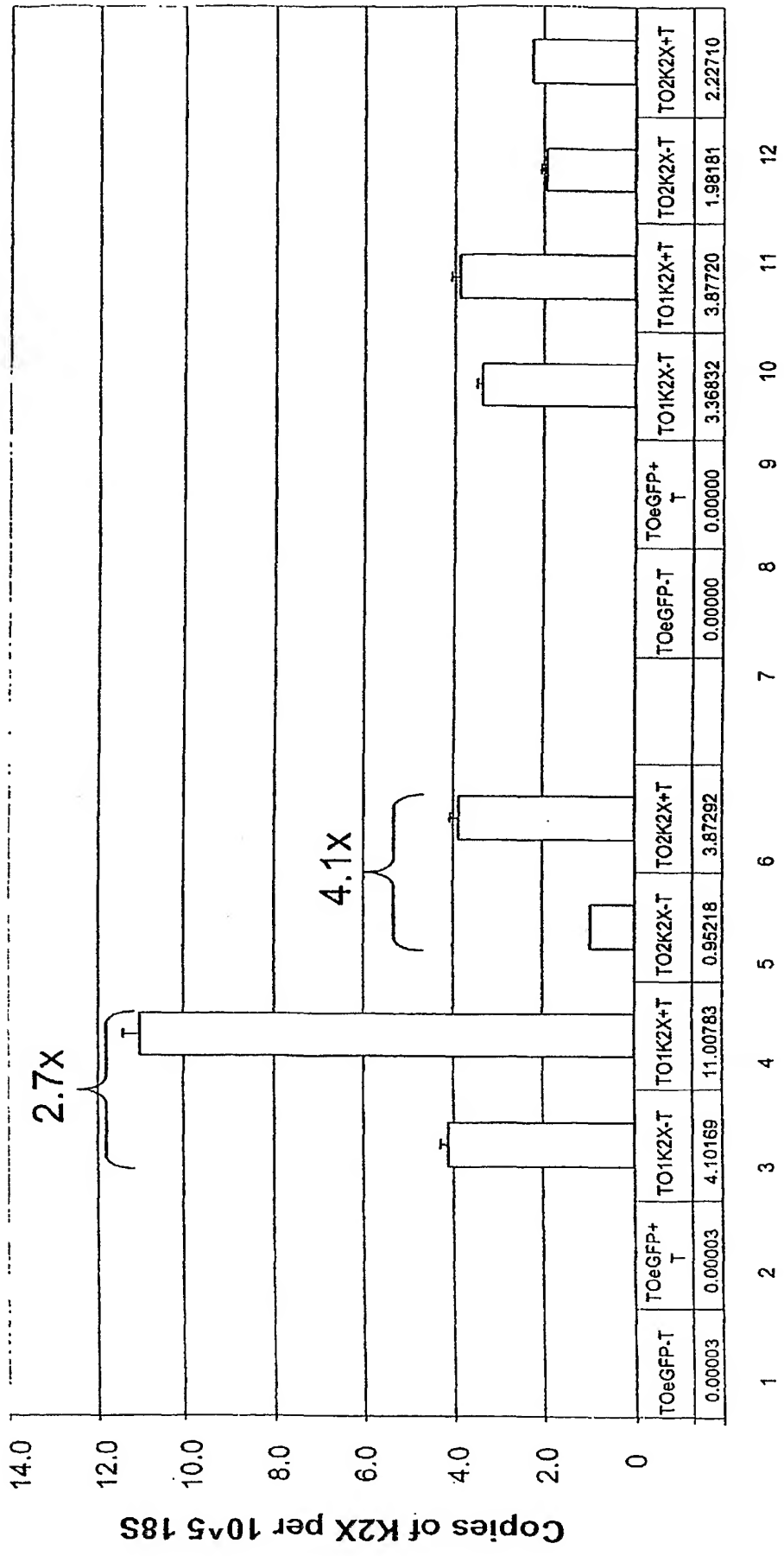

Relative K2X Expression in Transfected HeLa & HeLa::TR Cells
Copies of K2X per 10^5 18S
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0
2.7x
4.1x
TOeGFP-T
TOeGFP+T
TO1K2X-T
TO1K2X+T
TO2K2X-T
TO2K2X+T
TOeGFP-T
TOeGFP+T
TO1K2X-T
TO1K2X+T
TO2K2X-T
TO2K2X+T
0.00003
0.00003
4.10169
11.00783
0.95218
3.87292
0.00000
0.00000
3.36832
3.87720
1.98181
2.22710
1 2 3 4 5 6 7 8 9 10 11 12

Figure 11

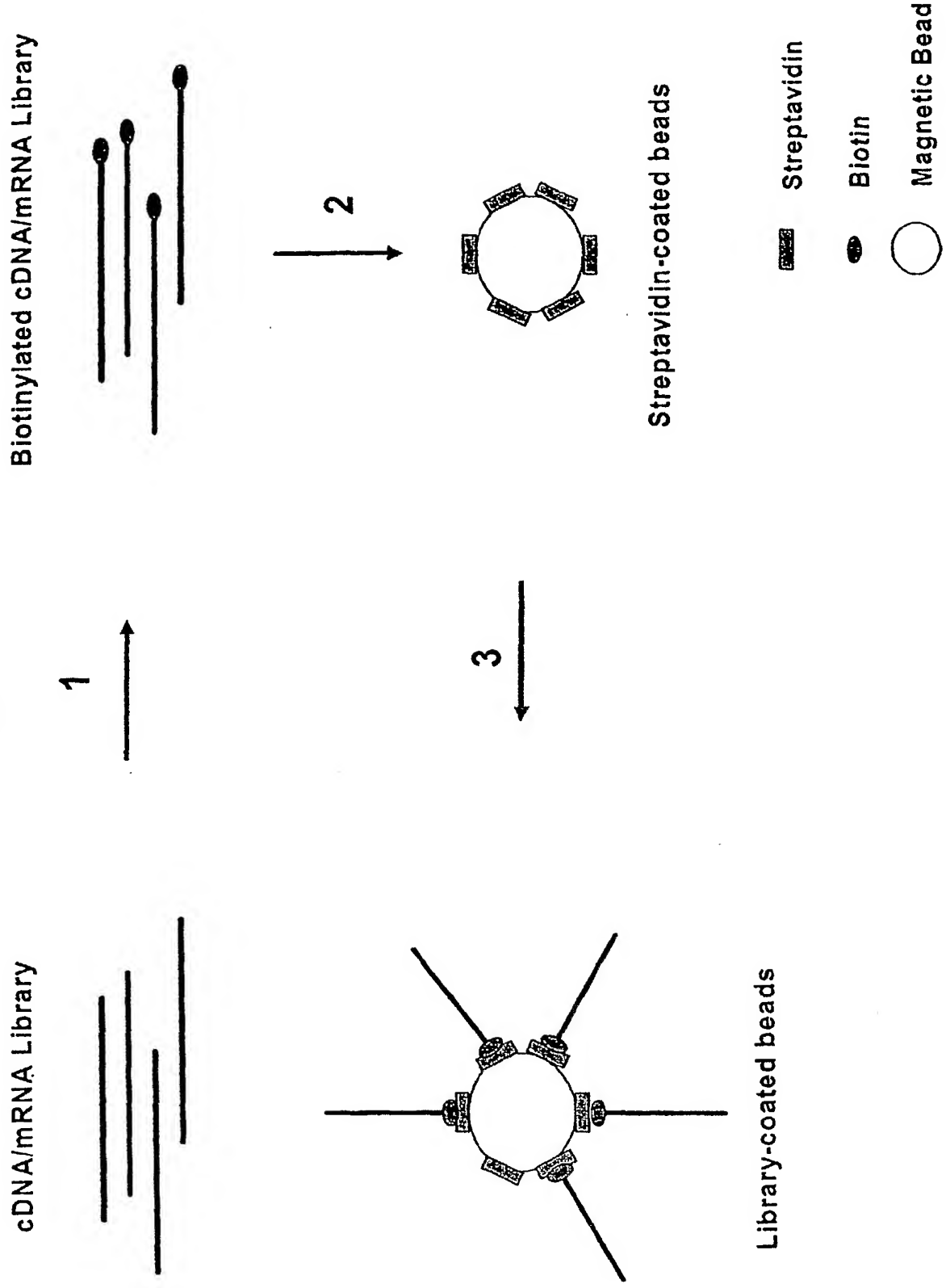

Biotinylated cDNA/mRNA Library
2
Streptavidin-coated beads
Streptavidin
Biotin
Magnetic Bead
1
3
cDNA/mRNA Library
Library-coated beads

Figure 12

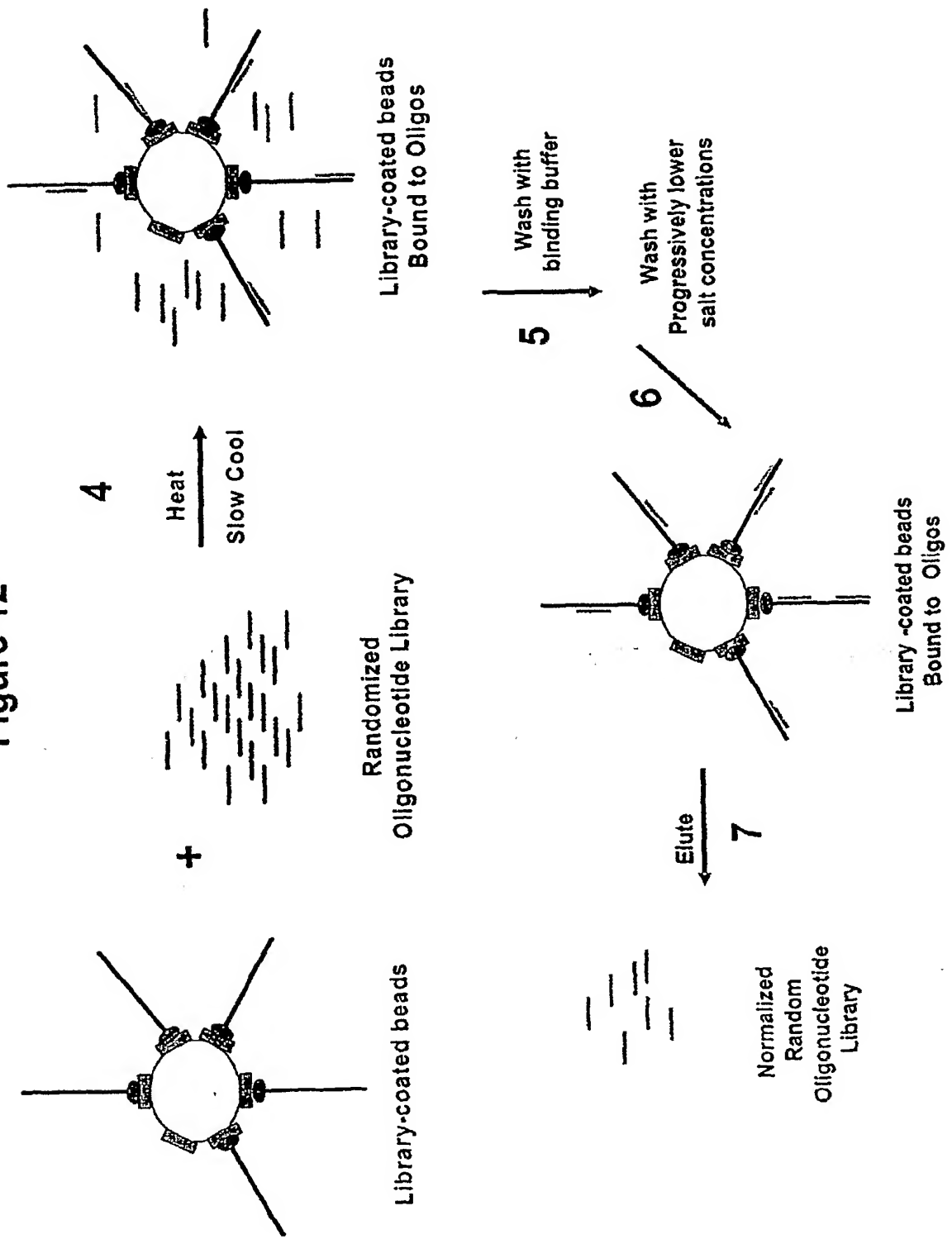

Library-coated beads
+
Randomized
Oligonucleotide Library
4
Heat
Slow Cool
Library-coated beads
Bound to Oligos
5
Wash with
binding buffer
6
Wash with
Progressively lower
salt concentrations
Elute
7
Normalized
Random
Oligonucleotide
Library
Library -coated beads
Bound to Oligos

Figure 13


A
B
C
D
Size
Specific 41mer
Specific 35mer
Non-specific 30mer
Non-specific 25mer
Specific 21mer